On Observational Variance Learning for Multivariate Bayesian Time Series and Related Models

Konstantinos Triantafyllopoulos

Thesis submitted for the degree of Doctor of Philosophy

Department of Statistics

University of Warwick

Coventry

CV4 7AL

February 2002

Στη μητέρα μου Τατιάνα και στον πατέρα μου Γιώργο

 Address for correspondence: 98 Pericleous Street, Halandri 15231, Athens, Greece, e-mail: `k_triantafyllopoulos@hotmail.com`

Στην αρχή το φως και η ώρα η πρώτη

που τα χείλη ακόμη στον πηλό

δοκιμάζουν τα πράγματα του κόσμου

ΑΥΤΟΣ

ο κόσμς ο μικρός, ο μέγας!

ΤΗ γλώσσα μου έδωσαν ελληνική·

το σπίτι φτωχικό στις αμμουδιές του Ομήρου.

Μονάχη έγνοια η γλώσσα μου στις αμμουδιές του Ομήρου.

Άξιον Εστί

Οδυσσέας Ελύτης

Contents

1 **Introduction** **1**

1.1 Introduction 1

1.2 Historical Review of Dynamic Models 1

1.3 Bayesian Forecasting 3

1.4 Plan of the Thesis 6

1.5 Notation 7

1.6 How to Read this Thesis 8

I Observational Variances **9**

2 **Introduction to the Normal DLM** **10**

2.1 Introduction 10

2.2 The Univariate Normal DLM 11

2.2.1 Definition of the Univariate Normal DLM 11

2.2.2 Observational Variances 12
2.3 The Multivariate Normal DLM 16
2.3.1 The General Multivariate Normal DLM 16
2.3.2 Observational Variances 19
2.4 Specification of the Evolution Variance 23
2.4.1 Single Discounting 23
2.4.2 Component Discounting 24
2.5 Superposition and Decomposition of DLMs 25
2.6 Observability and Model Design 27
2.6.1 Observability 28
2.6.2 Similar and Equivalent Models 31
2.6.3 Specification of $\boldsymbol{F}$ and $\boldsymbol{G}$ 34
2.7 Expert Intervention 39

3 A Regression DLM **42**
3.1 Introduction 42
3.2 The Static Regression Model 43
3.3 The Regression DLM 46
3.4 Relationship with the Common Components Model 51
3.5 Inversion Problems 53
3.6 The General Regression DLM 55

4 New Distributional Directions to the DLM **60**
4.1 Introduction 60
4.2 Generalized Wishart and Inverse Wishart Distributions 61
4.3 Generalized T Distribution 72

4.4 Matrix Normal DLMs 75

4.5 Retrospective Analysis 84

4.5.1 The ECCM Retrospective Parametric Distribution . . 84

4.5.2 Deleting Observations 86

4.5.3 Deleting Observations in the ECCM 89

4.6 Reference Analysis 92

4.7 An Approximation of the GMDLM 104

5 The General Multivariate DLM **109**

5.1 Introduction 109

5.2 Probability Modelling 110

5.2.1 The General Problem 110

5.2.2 Weak Posterior Prior Probability Modelling 111

5.3 The Estimate of the Observational Variance Matrix 116

5.4 Observational Variances 120

5.5 Limiting Results for Constant Observable Models 127

5.6 Illustration 130

6 Missing Observations **138**

6.1 Introduction 138

6.2 The Problem of Missing Observations 139

6.3 Missing Observations 143

6.3.1 The ECCM and the GMDLM 144

6.3.2 The Regression DLM 149

6.3.3 The General DLM of Chapter 5 150

6.4 Unequally Spaced Observations 154

6.5 Illustration 156

7 Time Varying Variances **175**

7.1 Introduction 175

7.2 Deterministic Variance Laws 176

7.2.1 The Linear Variance Law 176

7.2.2 General Deterministic Variance Laws 183

7.3 Stochastic Changes in Variance 188

7.3.1 The Extended Common Components Model 189

7.3.2 The General Multivariate DLM 195

7.4 Other Practical Variance Schemes 200

7.5 Illustration 202

8 Conclusions **207**

8.1 Introduction 207

8.2 Summary of Chapters 2-7 207

II Appendices **211**

A Linear Algebra **212**

A.1 Eigenvalues and Eigenvectors 212

A.2 Direct Product of Matrices, the Vec and Vech Operator 215

A.3 The Moore-Penrose Inverse 218

A.4 Linear Systems of Equations 220

A.5 The Limit of a Sequence of Matrices 221

A.6 Matrix Differentiation 225

A.7 The Powers of Symmetric Matrices 228

B Probability **232**

B.1 Moments of Matrix Distributions 232

B.2 The Matrix Normal Distribution 234

B.3 The Matrix T Distribution 235

C Tables **238**

C.1 Table of Chapter 5, Section 5.6 238

C.2 Tables of Chapter 6 244

C.2.1 Table of Section 6.3 244

C.2.2 Tables of Section 6.5 245

C.3 Table of Chapter 7, Section 7.5 265

List of Figures

4.1 Limiting behaviour of the expectation of an inverse Wishart random matrix 63

5.1 Bivariate simulated series: Series a 131

5.2 Variance estimation for $\sigma_{11} = 2$: Series a 132

5.3 Variance estimation for $\sigma_{12} = 1.5$: Series a 133

5.4 Variance estimation for $\sigma_{22} = 4$: Series a 133

5.5 Bivariate simulated series: Series b 134

5.6 Variance estimation for $\sigma_{11} = 2$: Series b 135

5.7 Variance estimation for $\sigma_{12} = 1.5$: Series b 136

5.8 Variance estimation for $\sigma_{22} = 4$: Series b 136

6.1 London Metal Exchange aluminium closing prices 159

6.2 Histogram and density estimation of the Cash Bid time series 161

6.3 Normal probability plot for the Cash Bid time series 162

6.4 One step forecasts vs actual values for the difference time series 163

6.5 One step errors for the difference time series 164

6.6 Normal probability plot excluding outliers 164

6.7 Comparison of the observational and evolution variances . . . 166

6.8 One step forecasts vs actual values for the difference time series after intervention 167

6.9 One step errors for the difference time series after intervention 167

6.10 Comparison of the correlation of the variables 171

6.11 Comparison of one step forecasts between the variables 173

6.12 Comparison of one step errors between the variables 174

7.1 Bivariate simulated series with time-varying observational variance 203

7.2 Discount factors and variance estimation for $\sigma_{11,t}$ 205

7.3 Discount factors and covariance estimation for $\sigma_{12,t}$ 206

7.4 Variance estimation for $\sigma_{22,t}$ 206

List of Tables

C.1 2 bivariate simulated series 238

C.2 Calculus of missing and observed values 244

C.3 London metal exchange official closing prices of aluminium in US $ per tonne 245

C.4 London metal exchange official closing prices of aluminium difference time series 258

C.5 Bivariate simulated series 265

Acknowledgements

To my supervisor P.J. Harrison I express my gratitude for his guidance and unceasing motivation during the last three years. Particularly I thank him for proofreading my thesis and making several invaluable suggestions.

I would also like to thank J.Q. Smith for his help and support having perfectly served as a co-supervisor at the last year of my studies.

I want to thank S. Ikonomakos, the director of the Commercial Department of Aluminium of Greece S.A.I.C., as well as A. Papageorgiou for providing me with the LME data of Section 6.5.

I am greatly indebted to my friend J. Pikoulas, with whom I attended conferences in Lisbon and Paris the last two years that provided me with a lot of motivation and confidence.

I owe thanks to all the staff in the Statistics Department at Warwick, especially to S.D. Jacka, J.B. Copas, and W.S. Kendall.

I would like to thank P.J. Matthews for her infinite patience as well as all my research fellows that contributed so much all these years to build a friendly and relaxing atmosphere. Particular thanks are due to B. Giglio, G. Montana, and F. Siannis.

I must thank the external examiner E.J. Godolphin and the internal examiner H.P. Wynn for several suggestions and improvements throughout the thesis.

I have to thank the Statistics Department for admitting me to graduate studying in Warwick to pursue studies in Statistics leading toward the degree of Doctor of Philosophy and for supporting me with a special research studentship.

Finally, I want to thank my parents and especially my father, Γιώργο, for teaching me the value of mathematics and for supporting me financially and mostly morally at good and bad times. It has always been an invaluable motivation to me our long discussions about science and its philosophy.

Declaration

I hereby declare that this thesis is my own work, except otherwise stated, and that it is not submitted or being under consideration for submission in any other university.

Chapter 3 is based on the paper Triantafyllopoulos and Pikoulas (2000), presented in the *20th International Symposium on Forecasting*, Lisbon 21-24 June 2000 (http://isf2000.deio.fc.ul.pt/ps.asp?Session=15). The mathematical part is entirely my own work. Following this we applied the method to more realistic intrusion detection systems to further illustrate our approach. The papers Triantafyllopoulos et al. (2000,2001) read in the *13th International Conference of Software and Systems Engineering and their Applications*, Paris 6-8 December 2000 and in the *8th International Confernece and Workshop on the Engineering of Computer Based Systems*, Washington 17-20 April 2001. The paper and the presentation of the last work are found in http://www.dcs.napier.ac.uk/~bill/ecbs_folder/ecbs_2001.pdf and http://www.dcs.napier.ac.uk/~bill/ecbs_folder/ecbs_main_colour_files/frame .htm.

The paper Triantafyllopoulos and Pikoulas (2000) has been submitted for publication to the *Journal of Forecasting.*

Early discussions with P.J. Harrison initiated the work for missing observations (Chapter 6). The relevant paper, Triantafyllopoulos (2000), provides the material for Chapters 4 and 6. This paper has been submitted for publication to the *Journal of Time Series Analysis.*

The paper Triantafyllopoulos and Harrison (2001) came naturally from unpublished notes of the latter author. The initial idea of the Weak Probability Modelling is not mine, but the formulation for the multivariate models and all the proofs in this paper and in Chapter 5 are my own work.

The neat results of the papers Triantafyllopoulos (2000) and Triantafyllopoulos and Harrison (2001) inspired me to provide for both these major classes of models a version for stochastic changes in the observational variance structure. This was completed in April 2001, while in Greece, and follows Triantafyllopoulos (2001).

The last two papers are considered for submission. All papers can be found in my personal webpage: http://www.stats.bris.ac.uk/~maxkt/

Summary

This thesis is concerned with variance learning in multivariate dynamic linear models (DLMs).

Three new models are developed in this thesis. The first one is a dynamic regression model with no distributional assumption of the unknown variance matrix. The second is an extension of a known model that enables comprehensive treatment of any missing observations. For this purpose new distributions that replace the inverse Wishart and matrix T and that allow conjugacy are introduced. The third model is the general multivariate DLM without any precise assumptions of the error sequences and of the unknown variance matrix. We find analytic updatings of the first two moments based on weak assumptions that are satisfied for the usual models.

Missing observations and time varying variances are considered in detail for every model. For the first time, deterministic and stochastic variance laws for the general multivariate DLM are presented. Also, by introducing a new

distribution that replaces the matrix-beta of a previous work, we prove results on stochastic changes in variance that are in line with missing observation analysis and variance intervention.

Notation

Symbol	Explanation
$\propto$	propotional
$\forall$	universal quantifier
$\exists$	existential quantifier
$\in$	member
$\bigcup_{i=1}^{n} A_i$	the union of the sets $A_1, \ldots, A_n$
$A \subset B$ $(B \supset A)$	the set A is a subset of B
$p(\cdot)$	density function
$\mathbb{N}, \mathbb{Z}, \mathbb{R}$	sets of natural, integer, and real numbers
$\mathbb{N}_n$	the excision of $\mathbb{N}$
$\mathbb{N}^*, \mathbb{N}_n^*, \mathbb{Z}^*, \mathbb{R}^*$	$\mathbb{N} - \{0\}, \mathbb{N}_n - \{0\}, \mathbb{Z} - \{0\}, \mathbb{R} - \{0\}$
$\mathbb{R}^m$	the set of all $m \times 1$ real vectors
$\mathbb{R}^{m \times n}$	the set of all $m \times n$ real matrices

continued on next page

continued from previous page

Symbol	Explanation
$[x]$	the integral part of the real number x
$\boldsymbol{X}$	vector or matrix
$\boldsymbol{X}'$	transpose of $\boldsymbol{X}$
$\boldsymbol{X}^{-1}$	inverse of $\boldsymbol{X}$
$\boldsymbol{X}^{-}$	generalized inverse of $\boldsymbol{X}$
$\boldsymbol{X}^{+}$	Moore-Penrose inverse of $\boldsymbol{X}$
$\boldsymbol{X}^{1/2}$	symmetric square root of $\boldsymbol{X}$
$\boldsymbol{X}^{-1/2}$	inverse of $\boldsymbol{X}^{1/2}$
$\lvert\boldsymbol{X}\rvert$	determinant of $\boldsymbol{X}$
$\lim_{t\to\infty} \boldsymbol{X}_t$	limit of a sequence of matrices
$\boldsymbol{N} \to \infty$	convergence of a diagonal matrix
$\Gamma_r(\cdot)$	the multivariate gamma function
$\beta_r(\cdot)$	the multivariate beta function
$\exp(\cdot)$	the exponential function
$\text{rank}(\cdot)$	the rank of a matrix
$\text{trace}(\cdot)$	the trace of a matrix
$\text{vec}(\cdot)$	vectorized form of an arbitrary matrix
$\text{vech}(\cdot)$	vectorized form of a symmetric matrix
$\otimes$	Kronecker product
$\lVert\boldsymbol{X}\rVert$	matrix norm
$\lVert\boldsymbol{X}\rVert_2$	the Eucledian norm
$\boldsymbol{F} : \mathcal{S} \to \mathcal{V}$	matrix valued function from $\mathcal{S}$ to $\mathcal{V}$

continued on next page

continued from previous page

Symbol	Explanation
$[\partial^i \boldsymbol{F}/\partial t^i]_{t=t_0}$	ith partial derivative of $\boldsymbol{F}$ at $t = t_0$
$J(\boldsymbol{X} \to \boldsymbol{F})$	Jacobian determinant of $\boldsymbol{F}$ in respect with $\boldsymbol{X}$
$\mathbf{1}_n$	the vector $(1, 0, \dots, 0)'$
$\boldsymbol{G}_n$	the duplication matrix
$\mathbf{0}$	zero-vector
$\mathbf{O}$	zero-matrix
$\boldsymbol{I}_n$	the $n \times n$ identity matrix
$\boldsymbol{I}$	identity matrix (when its dimension is clear)
$\boldsymbol{J}_n(\cdot)$	the $n \times n$ Jordan block matrix
$\boldsymbol{A}_{xy}$	regression matrix of $\boldsymbol{X}$ on $\boldsymbol{Y}$
$\perp_i$	ith order independence ($i = 1, 2$)
$\perp\!\!\!\perp_i$	ith order mutual independence ($i = 1, 2$)
$\perp\!\!\!\perp$	mutual independence
$(\mathcal{U})$	uncorrelation property
$\mathrm{E}[\cdot]$	expectation of a random vector or a random matrix
$\mathrm{V}[\cdot]$	variance of a random vector
$\mathrm{C}[\cdot, \cdot]$	covariance of two random vectors
$\mathrm{V}_l[\cdot]$	left variance of a random matrix
$\mathrm{V}_r[\cdot]$	right variance of a random matrix
$\boldsymbol{X} \sim [\boldsymbol{M}, \boldsymbol{C}, \boldsymbol{S}]$	abbreviation of $\mathrm{E}[\boldsymbol{X}] = \boldsymbol{M}$, $\mathrm{V}_l[\boldsymbol{X}] = \boldsymbol{C}$, $\mathrm{V}_r[\boldsymbol{X}] = \boldsymbol{S}$
$\mathrm{N}[\boldsymbol{M}, \boldsymbol{C}, \boldsymbol{S}]$	matrix normal distribution
$\mathrm{T}_n[\boldsymbol{M}, \boldsymbol{C}, \boldsymbol{S}]$	matrix T distribution

continued on next page

continued from previous page	
Symbol	Explanation
$\mathrm{G}[a, b]$	gamma distribution
$\mathrm{W}_r[k, \boldsymbol{R}]$	Wishart distribution
$\mathrm{W}_r^{-1}[k, \boldsymbol{R}]$	inverse Wishart distribution
$\mathrm{Beta}_r[a, b]$	matrix beta distribution
$\mathrm{GW}[\boldsymbol{S}, \boldsymbol{N}, m]$	general Wishart distribution
$\mathrm{GW}^{-1}[\boldsymbol{S}, \boldsymbol{N}, m]$	general inverse Wishart distribution
$\mathrm{NGW}^{-1}[\boldsymbol{M}, \boldsymbol{C}, \boldsymbol{S}, \boldsymbol{N}, m]$	joint normal general inverse Wishart distribution
$\mathrm{GT}[\boldsymbol{m}, \boldsymbol{C}, \boldsymbol{S}, \boldsymbol{N}, m]$	general T distribution
$\mathrm{GB}[\frac{1}{2}\boldsymbol{N}_1, \frac{1}{2}\boldsymbol{N}_2]$	general beta distribution
D_t	information set at time t
$\{\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_t, \boldsymbol{W}_t\}$	quadruple that defines a dynamic linear model
$\{\boldsymbol{F}_t, \boldsymbol{G}_t, \cdot, \cdot\}$	same as above
CCM	Common Components Model
DLM	Dynamic Linear Model
ECCM	Extended Common Components Model
GMDLM	General Multivariate Dynamic Linear Model
LME	London Metal Exchange
NDRDLM	Normal Discount Regression Dynamic Linear Model
SPD	Symmetric Positive Definite (matrix)

CHAPTER 1

Introduction

1.1 Introduction

This chapter introduces Bayesian forecasting as well as giving an account of how this thesis is organised. Section 1.2 gives the background of Bayesian forecasting. Section 1.3 presents the general framework under which Bayesian forecasting and dynamic models operate. The last three sections deal with the organisation of the thesis.

1.2 Historical Review of Dynamic Models

Bayesian forecasting and dynamic models have a history of almost half a century. This goes back to the late 1950s in short-term forecasting with the works by Harrison [15, 16] among other authors like Box, Brown, Cox, Holt, Muth, and Whittle. Relevant references can be found in [15, 16]. At almost the same time similar developments arose in the areas of systems and control engineering by Kalman [26, 27]. In the 1970s dynamic modelling admitted significant developments with [18, 19]. The reference [19] is the formal introduction of dynamic models and the feedback of a number of discussants is particularly useful. Although sometimes controversial, depending on the point of view that every discussant has about forecasting, Bayesian forecasting was at least promising in the late 1970s.

In the 1980s and 1990s there was a rapid development in the area including [1, 2, 3, 4, 7, 12, 21, 17, 20, 22, 23, 30, 33, 34, 35, 36, 37, 48]. Two main texts appeared the last decade. [31] gives an introduction to Bayesian forecasting from a practical point of view. [51] is a comprehensive account of dynamic models and covers areas such as univariate DLMs (dynamic linear models), model design, intervention and monitoring, dynamic regression, model irregularities, multi-process models, non-linear dynamic models, generalised dynamic linear models, MCMC methods in dynamic models, and multivariate DLMs.

Multivariate dynamic models appeared initially in [19]. However, it is clear that the focus by then was on univariate modelling. Important contributions to multivariate dynamic modelling include [3, 4, 33, 34, 35, 36].

Today, multivariate modelling admits a significant area of Bayesian forecasting and in the future it is likely to receive even more attention. It is worthwhile saying that there is an interrelation between Bayesian forecasting and other areas of Bayesian methods, like Bayesian networks and graphical models. These areas as well as other areas in statistics like simulation methods for dynamic models, are expected to have a significant growth in the future. In [50] there is a description of dynamic modelling with useful trends and possible future research areas. The interested reader can find there an extended reference list. From the Kalman filtering point of view, [10] is an excellent reference which provides both a frequentist and Bayesian treatment of state space methods.

1.3 Bayesian Forecasting

Forecasting systems are integrated systems. That is, they are systems that try to predict a phenomenon that does not operate within a certain law.

A typical forecasting system comprises various work elements. One of them is statistical modelling. Yet it is not the only one. Our systems work with the principle of *Management by exception.* That means that our model will be appropriate for forecasting unless special circumstances arise. Often there is external qualitative information affecting the observations of interest. Thus, forecasting has to be adjusted to this information and merely the past observations are not appropriate for prediction, in this case. For example a change in a government's taxation plan may affect the price plans of a company. Routine forecasting is proved poor if it is not able to take into account such external information. This can be restored and evaluated by the company's experts that are not statisticians. In short-term forecasting fast response is a necessity. A team formed by the company's experts and statisticians can be very effective. In general, the most difficult thing is to evaluate the impact of qualitative information and how this can be incorporated in the models. Such systems are called open because they take into account external subjective information. We have found that Bayesian statistics offers a unique framework for handling such information. In this thesis we stick to the mathematical interpretation of the forecasting system, although we shall always bear in mind that this is only a part of the complete forecasting system.

Dynamic modelling operates on discrete time and especially on a subset of the set of integer numbers. In this thesis it is assumed that observations are collected in equal intervals. However, in Chapter 6 the problem of unequally spaced observations is briefly discussed.

Suppose that at time t the $r \times 1$ observation vector of interest is denoted by $\boldsymbol{Y}_t$. $\boldsymbol{\theta}_t$ is an $n \times 1$ vector of states that evolves stochastically. This is what makes the model dynamic. The model may be expressed generally by

$$p(\boldsymbol{Y}_t | \boldsymbol{\theta}_t, \boldsymbol{Y}_{t-1}, \ldots, \boldsymbol{Y}_1, D_0), \tag{1.1}$$

$$p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \boldsymbol{Y}_{t-1}, \ldots, \boldsymbol{Y}_1, D_0), \tag{1.1$'$}$$

where the above densities are assumed known and the conditioning vectors $\boldsymbol{Y}_{t-1}, \ldots, \boldsymbol{Y}_1$ are not any more random quantities since they are all known at time t. To make this clear, the information set, D_t, is defined as

$$D_t = \{D_{t-1}, \boldsymbol{Y}_t\} = \{D_0, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t\},$$

for $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t$ observed and D_0 is the initial information.

Also the initial density of $(\boldsymbol{\theta}_0 | D_0)$,

$$p(\boldsymbol{\theta}_0 | D_0), \tag{1.2}$$

is assumed known.

Suppose now that at time $t-1$ given D_{t-1}, the density of $p(\boldsymbol{\theta}_{t-1}|D_{t-1})$ is known. The prior distribution of the states at time t is calculable from (1.1′) and the joint distribution of $\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}_{t-1}$ given D_{t-1} as

$$p(\boldsymbol{\theta}_t|D_{t-1}) = \int p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, D_{t-1})p(\boldsymbol{\theta}_{t-1}|D_{t-1})\, d\boldsymbol{\theta}_{t-1}, \tag{1.3}$$

where the integrand term is integrated over $\mathbb{R}^n$.

Then the one-step forecast distribution is calculated by

$$p(\boldsymbol{Y}_t|D_{t-1}) = \int p(\boldsymbol{Y}_t|\boldsymbol{\theta}_t, D_{t-1})p(\boldsymbol{\theta}_t|D_{t-1})\, d\boldsymbol{\theta}_t, \tag{1.4}$$

where the integrand term is integrated over $\mathbb{R}^n$.

After this the posterior is updated from time $t-1$ to t, when D_t becomes available according to the Bayes' Theorem as

$$p(\boldsymbol{\theta}_t|D_t) \propto p(\boldsymbol{\theta}_t|D_{t-1})p(\boldsymbol{Y}_t|\boldsymbol{\theta}_t, D_{t-1})$$

and the cycle starts anew.

The quantity $p(\boldsymbol{Y}_t|\boldsymbol{\theta}_t, D_{t-1})$ of the last equation is usually the likelihood function of $\boldsymbol{\theta}_t$ given the value $\boldsymbol{Y}_t$ at time t. Also the density (1.2) allows for a complete updating from time 1 to any time t.

In general the integrals of equations (1.3), (1.4) will not be analytically calculated. In such cases simulation based methods may be appropriate. However, when the model is linear and the errors are independently normally distributed there are analytical forms and the updatings coincide with those derived by Kalman, see [26, 27]. Note that when there are other unknown quantities like error variances and they are viewed as random, integration in equations (1.3), (1.4) will be in addition with respect to these quantities. This makes the case of analytic calculations more difficult.

1.4 Plan of the Thesis

The aim of this thesis is first to provide analytic results to the problem of multivariate dynamic linear modelling and forecasting with an unknown observational variance matrix and second to provide a general treatment of the multivariate methods mainly as far as model specification and design is concerned. The thesis is organised as follows.

Chapter 2 is an introduction to the multivariate dynamic linear model. It presents a unified approach to multivariate methods avoiding the reference to the univariate methods. Model design and specification receive special attention. Also intervention aspects are discussed.

Chapter 3 presents a new regression dynamic model. Its relationship with current models and extensions to the matrix-variate case are discussed.

Chapter 4 is central to the thesis. Introducing new distributions, overcomes problems of current models related to missing observations and intervention. This chapter discusses the new distributions with their basic properties and develops the new models with these distributions.

Chapter 5 defines and develops a new perspective to dynamic modelling. According to this approach modelling is possible for any linear model without a precise specification of the distributions of the error terms as long as some weak assumptions are satisfied. The general problem of observational variances for the first time finds analytical formulae for the first two moments. The first two moments of the filtering distribution are derived as well as some limiting results. Simulations show the capabilities of the new method against existing ones.

Chapter 6 deals with incomplete data in multivariate DLMs. Results from Chapter 4 are applied in this chapter to overcome missing observation problems. Also methods developed in Chapters 3, 5 are modified for this reason. The problem of unequally spaced observations is discussed. An application with real data illustrates the methods.

Chapter 7 is a treatment of time varying variances. It includes deterministic variance laws and stochastic changes in variance. For the latter the new distributions defined in Chapter 4 and another new introduced in this chapter generalize the existing methods allowing missing observations and variance intervention, as well as being more sensible than older models. A methodology for the general model of Chapter 5 is developed. Implementation aspects are discussed via a simulation.

1.5 Notation

The notation of this thesis follows West and Harrison, [51]. An attempt was made to keep the terminology as consistent as possible, being in line with the above reference. Known material is presented exactly as it appears in [51], without the proofs. Only the proofs of new results are presented.

Theorems and equations are numbered within each of the chapters, while sections follow the usual sub-numbering in every chapter. So the notation Section 2.5.2 refers to Subsection 2 of Section 5 of Chapter 2. Similarly, the notation Theorem 2.10 refers to the Theorem 10 of Chapter 2.

The various integrals are Lebesgue integrals and the respective integration set is usually omitted, where this is clear from the context.

1.6 How to Read this Thesis

The chapters of this thesis have a sequential development and it is recommended that they be read as ordered. However, some alternative reading could be the following. After Chapters 1, 2, Chapters 3, 4, 5 follow in any order and then again Chapters 6, 7 in any order with low loss. The reader might also obtain an overview of the thesis by first reading the last chapter giving conclusions.

Part I

Observational Variances

CHAPTER 2

Introduction to the Normal DLM

2.1 Introduction

In this chapter an introduction to the main multivariate normal DLM methodology is explored. For notational simplicity this is just referred to as DLM. The content of the chapter can be characterised as a review of existing methods. The material is not new, unless otherwise stated.

In Section 2.2 the univariate DLM is briefly discussed. Multivariate DLMs are described in Section 2.3 with the focus on observational variances. The last three sections deal with model specification and design as well as possible actions of model assessment and improvement. Section 2.4 discusses the specification of the evolution variance matrix. The next section discusses the superposition and decomposition principle of DLMs. In Section 2.6 aspects of model design are introduced, some for the first time for the multivariate case. Definitions 2.1, 2.7 and Theorems 2.7, 2.8, 2.9, are new and provide the complete analogue to the univariate case. Finally, Section 2.7 explores possible modes of expert intervention.

2.2 The Univariate Normal DLM

In this section the univariate DLM is briefly described for comparison with the multivariate model as well as for completion purposes. The main references are [19] and [51].

2.2.1 Definition of the Univariate Normal DLM

The general univariate model is given by

$$Y_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, \qquad \nu_t \sim \mathrm{N}[0, V_t], \tag{2.1}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{W}_t], \tag{2.1$'$}$$

where Y_t is a scalar observation series, $\boldsymbol{F}_t$ a known $n \times 1$ design vector, $\boldsymbol{\theta}_t$ an $n \times 1$ state vector, ν_t a normal random variable, V_t a variance, $\boldsymbol{G}_t$ a known $n \times n$ evolution matrix, $\boldsymbol{\omega}_t$ an $n \times 1$ normal random vector, and $\boldsymbol{W}_t$ an $n \times n$ variance matrix. Every time that an observation is received the information we get is richer. This is expressed by the information set $D_t = \{D_{t-1}, Y_t\}$. This simply says that prior to t the information we have is D_{t-1}, and it is this information we are going to use to forecast until the value of Y_t is available. Initially we start with a D_0, which may be the null set, if no such information exists. Equation (2.1) is called the **observation equation** and equation (2.1$'$) the **evolution or system equation**. The former is the series of interest either for forecasting or for analysis and control. The latter is a first order Markov chain that expresses the possible evolution, both stochastic and in respect to t, of the parameters or states $\boldsymbol{\theta}_t$.

Alternatively equations (2.1), (2.1′) can be written as

$$(Y_t|\boldsymbol{\theta}_t, D_{t-1}) \sim \mathrm{N}[\boldsymbol{F}_t'\boldsymbol{\theta}_t, V_t],$$

$$(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, D_{t-1}) \sim \mathrm{N}[\boldsymbol{G}_t\boldsymbol{\theta}_t, \boldsymbol{W}_t].$$

Conditional independence applies, see [51, pages 98,136]. According to this at time t given D_{t-1}, $\boldsymbol{\theta}_t$ is independent of all the previous states $\boldsymbol{\theta}_{t-i}$, $(1 \leq i \leq t)$ and all the future ones, $\boldsymbol{\theta}_{t+j}$, $(j > 0)$. In other words given the present, the past is independent of the future . Moreover, the distribution of Y_t is explicitly known when $\boldsymbol{\theta}_t$ and D_{t-1} are known. This is given by the first of the above equations.

The choice of the variances V_t and $\boldsymbol{W}_t$ is critical for design and implementation purposes. For example if the elements of $\boldsymbol{W}_t$ are allowed to be very large the model will be very unstable with very low precision. On the other extreme if they are very low, the model will be almost static, assuming that there is little stochastic evolution. Things are even more difficult for the modeller considering the fact that $\boldsymbol{\theta}_t$ are unobservable quantities, thus $\boldsymbol{W}_t$ has to be chosen in accord with the system components that somehow are related with the actual series Y_t and the information sets. At the moment the variances $\boldsymbol{W}_t$ are assumed known.

2.2.2 Observational Variances

Practitioners find particular difficulty in specifying the variance V_t, even when it is constant over time, $V_t = V$. The Bayesian paradigm offers a very flexible way to handle with this problem, specifying V as a random variable having a skewed inverse-gamma distribution. For the rest of the thesis we use the Greek letter Σ instead of V to signify that this variance is unknown. In order to proceed analytically a scaled version of the model (2.1), (2.1′) is needed. So the working model is

$$Y_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, \qquad \nu_t \sim \mathrm{N}[0, \Sigma], \tag{2.2}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathrm{N}[\mathbf{0}, \Sigma\boldsymbol{W}_t^*], \tag{2.2′}$$

with the initial distributions

$$(\boldsymbol{\theta}_0|D_0, \phi) \sim \mathrm{N}[\boldsymbol{m}_0, \Sigma\boldsymbol{C}_0^*],$$

$$(\phi|D_0) \sim \mathrm{G}[n_0/2, n_0S_0/2],$$

for some known initial quantities $\boldsymbol{m}_0$, $\boldsymbol{C}_0^*$, n_0, and S_0, where $\phi = \Sigma^{-1}$.

Theorem 2.1. *With the above DLM, the following distributional results obtain at each time $t \geq 1$.*

(a) Conditional on Σ:

$$(\boldsymbol{\theta}_{t-1}|D_{t-1}, \Sigma) \sim N[\boldsymbol{m}_{t-1}, \Sigma\boldsymbol{C}_{t-1}^*],$$

$$(\boldsymbol{\theta}_t|D_{t-1}, \Sigma) \sim N[\boldsymbol{a}_t, \Sigma\boldsymbol{R}_t^*],$$

$$(Y_t|D_{t-1}, \Sigma) \sim N[f_t, \Sigma Q_t^*],$$

$$(\boldsymbol{\theta}_t|D_t, \Sigma) \sim N[\boldsymbol{m}_t, \Sigma\boldsymbol{C}_t],$$

with

$$\begin{aligned} \boldsymbol{a}_t &= \boldsymbol{G}_t \boldsymbol{m}_{t-1}, & \boldsymbol{R}_t^* &= \boldsymbol{G}_t \boldsymbol{C}_{t-1}^* \boldsymbol{G}_t' + \boldsymbol{W}_t^*, \\ f_t &= \boldsymbol{F}_t' \boldsymbol{a}_t, & Q_t^* &= \boldsymbol{F}_t' \boldsymbol{R}_t^* \boldsymbol{F}_t + 1, \\ e_t &= Y_t - f_t, & \boldsymbol{A}_t &= \boldsymbol{R}_t^* \boldsymbol{F}_t / Q_t^*, \\ \boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{A}_t e_t, \quad \textit{and} & \boldsymbol{C}_t^* &= \boldsymbol{R}_t^* - \boldsymbol{A}_t \boldsymbol{A}_t' Q_t^*. \end{aligned}$$

(b) For the precision $\phi = \Sigma^{-1}$:

$$(\phi | D_{t-1}) \sim G\left[\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2}\right],$$
$$(\phi | D_t) \sim G\left[\frac{n_t}{2}, \frac{n_t S_t}{2}\right],$$

where $n_t = n_{t-1} + 1$ and $n_t S_t = n_{t-1} S_{t-1} + e_t^2 / Q_t^*$.

(c) Unconditional on Σ:

$$(\boldsymbol{\theta}_{t-1} | D_{t-1}) \sim T_{n_{t-1}}[\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}],$$
$$(\boldsymbol{\theta}_t | D_{t-1}) \sim T_{n_{t-1}}[\boldsymbol{a}_t, \boldsymbol{R}_t],$$
$$(Y_t | D_{t-1}) \sim T_{n_{t-1}}[f_t, Q_t],$$
$$(\boldsymbol{\theta}_t | D_t) \sim T_{n_t}[\boldsymbol{m}_t, \boldsymbol{C}_t],$$

where $\boldsymbol{R}_t = S_{t-1}\boldsymbol{R}_t^*$, $Q_t = S_{t-1} Q_t^*$, and $\boldsymbol{C}_t = S_t \boldsymbol{C}_t^*$.

(d) Operational definitions of updating equations:

$$\begin{aligned} \boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{A}_t e_t', & \boldsymbol{C}_t &= \frac{S_t}{S_{t-1}}(\boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{A}_t' Q_t), \\ n_t &= n_{t-1} + 1, & S_t &= S_{t-1} + \frac{S_{t-1}}{n_t}\left(\frac{e_t^2}{Q_t} - 1\right), \\ Q_t &= \boldsymbol{F}_t' \boldsymbol{R}_t \boldsymbol{F}_t + S_{t-1}, \quad \textit{and} & \boldsymbol{A}_t &= \boldsymbol{R}_t \boldsymbol{F}_t / Q_t. \end{aligned}$$

Proof. See [51, chapter 4]. □

The above theorem provides the analysis for both known and unknown observational variances. Using the identities of the Gamma distribution we have $\mathrm{E}[\Sigma|D_t] = n_t S_t/(n_t - 2)$ and $\mathrm{E}[\Sigma^{-1}|D_t] = S_t^{-1}$. Thus, as $t \to \infty$, Σ concentrates about its mode, asymptotically degenerating. The above analysis is not restricted to the scaled model. It is generally applied to the model (2.1), (2.1′) when $V_t = \Sigma$ is unknown (see [51, chapter 4]).

Identities

(1) $Q_t = (1 - \boldsymbol{F}_t'\boldsymbol{A}_t)^{-1} S_{t-1}$;

(2) $\boldsymbol{A}_t = \boldsymbol{C}_t \boldsymbol{F}_t / S_t$;

(3) $\boldsymbol{C}_t^{-1} \approx \boldsymbol{R}_t^{-1} + \boldsymbol{F}_t \boldsymbol{F}_t' / S_t$.

Proof. (1) From the updating $\boldsymbol{A}_t = \boldsymbol{R}_t \boldsymbol{F}_t / Q_t$ we have $1 - \boldsymbol{F}_t'\boldsymbol{A}_t = S_{t-1} Q_t^{-1}$, from which (1) is directly derived. (2) First note that

$$\boldsymbol{R}_t \boldsymbol{F}_t \boldsymbol{F}_t' \boldsymbol{A}_t / S_{t-1} = \boldsymbol{A}_t \boldsymbol{A}_t' \boldsymbol{F}_t Q_t / S_{t-1}. \tag{2.3}$$

Then, the updatings of $\boldsymbol{C}_t$ and $\boldsymbol{A}_t$ of Theorem 2.1 imply

$$\boldsymbol{A}_t = \boldsymbol{C}_t \boldsymbol{F}_t / S_t - \boldsymbol{C}_t \boldsymbol{F}_t \boldsymbol{F}_t' \boldsymbol{A}_t / S_t + \boldsymbol{A}_t \boldsymbol{A}_t' \boldsymbol{F}_t Q_t / S_{t-1} - \boldsymbol{A}_t \boldsymbol{A}_t' \boldsymbol{F}_t \boldsymbol{F}_t' \boldsymbol{A}_t Q_t / S_{t-1}$$

and using equation (2.3) we have (2). (3) Using the updatings of $\boldsymbol{C}_t$ and $\boldsymbol{A}_t$ of Theorem 2.1 together with (2)

$$\begin{aligned} (\boldsymbol{R}_t^{-1} + \boldsymbol{F}_t \boldsymbol{F}_t' / S_t) \boldsymbol{C}_t &= \frac{S_t}{S_{t-1}} (\boldsymbol{I} - \boldsymbol{F}_t \boldsymbol{A}_t') + \boldsymbol{F}_t \boldsymbol{F}_t' \boldsymbol{C}_t / S_t \\ &\approx \boldsymbol{I}. \end{aligned}$$

Similarly it is shown that $\boldsymbol{C}_t(\boldsymbol{R}_t^{-1} + \boldsymbol{F}_t \boldsymbol{F}_t'/S_t) \approx \boldsymbol{I}$, and the result follows. □

As a special case of model (2.1), (2.1′) we consider the so called constant first-order polynomial model with known observational variances . That is, model (2.1), (2.1′) with $n = 1$, $F_t = 1$, $G_t = 1$, V known, and $W_t = W$. So the model is

$$Y_t = \mu_t + \nu_t, \qquad \nu_t \sim \mathrm{N}[0, V], \tag{2.4}$$

$$\mu_t = \mu_{t-1} + \omega_t, \qquad \omega_t \sim \mathrm{N}[0, W]. \tag{2.4′}$$

Although a very simple model, this is a model of noticeable interest. The following limiting results are important as they give an aid to specifying W.

Theorem 2.2. *Define $p = W/V$. As $t \to \infty$, $A_t \to A$, $C_t \to C = AV$, where*

$$A = \frac{p}{2}\left(\sqrt{1+\frac{4}{p}} - 1\right) = \frac{2}{1+\sqrt{1+\frac{4}{p}}}.$$

Proof. See [51, chapter 2]. □

2.3 The Multivariate Normal DLM

2.3.1 The General Multivariate Normal DLM

Multivariate dynamic models are the physical extension of the univariate models to several series. The multivariate analogue to model (2.1), (2.1′) is

$$\boldsymbol{Y}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{V}_t], \tag{2.5}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{W}_t], \tag{2.5′}$$

where $\boldsymbol{Y}_t$ is an $r \times 1$ observation vector, $\boldsymbol{F}_t$ a known $n \times r$ design matrix, $\boldsymbol{\theta}_t$ an $n \times 1$ state vector, $\boldsymbol{\nu}_t$ a normal $r \times 1$ random vector, $\boldsymbol{V}_t$ an $r \times r$ variance matrix, G_t a known $n \times n$ evolution matrix, ω_t an $n \times 1$ normal random vector, and W_t an $n \times n$ variance matrix. This model is often denoted by the quadruple

$$\{F_t, G_t, V_t, W_t\}.$$

The information set at time t is defined as $D_t = \{D_{t-1}, Y_t\}$ and similar comments apply as in the univariate case. Also, the initial distribution for $(\theta_0|D_0)$ is

$$(\theta_0|D_0) \sim \mathrm{N}[m_0, C_0],$$

for some known quantities m_0 and C_0.

If V_t, W_t are known there is a very similar analysis to the univariate case, presented in the following theorem.

Theorem 2.3. *One-step forecast and posterior distributions in the model just defined are given, for each t, as follows.*

(a) Posterior at $t-1$:

$$(\theta_{t-1}|D_{t-1}) \sim N[m_{t-1}, C_{t-1}],$$

for some known quantities m_{t-1}, C_{t-1}.

(b) Prior at t:

$$(\theta_t|D_{t-1}) \sim N[a_t, R_t],$$

where

$$a_t = G_t m_{t-1} \quad \text{and} \quad R_t = G_t C_{t-1} G_t' + W_t.$$

(c) One-step forecast:

$$(Y_t|D_{t-1}) \sim N[f_t, Q_t],$$

where

$$\boldsymbol{f}_t = \boldsymbol{F}_t'\boldsymbol{a}_t \qquad \textit{and} \qquad \boldsymbol{Q}_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + \boldsymbol{V}_t.$$

(d) Posterior at t:

$$(\boldsymbol{\theta}_t|D_t) \sim N[\boldsymbol{m}_t, \boldsymbol{C}_t],$$

with

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t\boldsymbol{e}_t \qquad \textit{and} \qquad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{Q}_t\boldsymbol{A}_t',$$

where

$$\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{Q}_t^{-1} \qquad \textit{and} \qquad \boldsymbol{e}_t = \boldsymbol{Y}_t - \boldsymbol{f}_t.$$

Proof. See [51, chapter 16]. □

The filtered marginal distributions are defined by

$$(\boldsymbol{\theta}_{t-k}|D_t) \sim \mathrm{N}[\boldsymbol{a}_t(-k), \boldsymbol{R}_t(-k)], \tag{2.6}$$

where

$$\boldsymbol{a}_t(-k) = \boldsymbol{m}_{t-k} + \boldsymbol{B}_{t-k}[\boldsymbol{a}_t(-k+1) - \boldsymbol{a}_{t-k+1}],$$
$$\boldsymbol{R}_t(-k) = \boldsymbol{C}_{t-k} + \boldsymbol{B}_{t-k}[\boldsymbol{R}_t(-k+1) - \boldsymbol{R}_{t-k+1}]\boldsymbol{B}_{t-k}',$$

with $\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{G}_{t+1}'\boldsymbol{R}_{t+1}^{-1}$ and starting values $\boldsymbol{a}_t(0) = \boldsymbol{m}_t$, $\boldsymbol{R}_t(0) = \boldsymbol{C}_t$, together with $\boldsymbol{a}_{t-k}(1) = \boldsymbol{a}_{t-k+1}$, and $\boldsymbol{R}_{t-k}(1) = \boldsymbol{R}_{t-k+1}$.

Model design is related to the specification of $\boldsymbol{V}_t$, $\boldsymbol{W}_t$, $\boldsymbol{F}_t$, and $\boldsymbol{G}_t$. Often $\boldsymbol{W}_t$ is specified via the discounting approach (see Section 2.4). The choice of $\boldsymbol{F}_t$ and $\boldsymbol{G}_t$ is discussed in Section 2.6. So it remains the specification of $\boldsymbol{V}_t$. Again it is used a constant $\boldsymbol{V}_t = \boldsymbol{V} = \boldsymbol{\Sigma}$ to signify that the observational variance matrix is unknown.

2.3.2 Observational Variances

The problem of unknown variances is far from an easy task. Notice that there are $r(r+1)/2$ unknown elements of $\boldsymbol{\Sigma}$ to be specified. Many practitioners assume that all the covariances of $\boldsymbol{\Sigma}$ are equal to zero, or to a constant c. Thus, the problem is reduced to specifying r, or $r+1$ unknown parameters respectively. Then, instead of employing the model (2.5), (2.5′) they use r univariate DLMs. However, such assumptions are unwarranted. The problem of estimating the different $r(r+1)/2$ elements of $\boldsymbol{\Sigma}$ is even bigger, considering the computational difficulties of any numerical method, as r increases. The natural extension of the univariate case is to model $\boldsymbol{\Sigma}$ with an inverse Wishart distribution. This approach, unfortunately, lacks conjugacy. The resulting distributions of $(\boldsymbol{\theta}_{t-1}|D_{t-1})$, $(\boldsymbol{\theta}_t|D_{t-1})$, $(\boldsymbol{Y}_t|D_{t-1})$, and $(\boldsymbol{\theta}_t|D_t)$, unconditional of $\boldsymbol{\Sigma}$, are not tractable. Quintana ([33],[34]) developed a matrix-variate DLM, known as CCM (Common Components Model), using matrix normal and T distributions rather than multivariate ones. This is described below.

Suppose we have r univariate DLMs $\{\boldsymbol{F}_t, \boldsymbol{G}_t, V_t\sigma_j^2, \boldsymbol{W}_t\sigma_j^2\}$, $\forall\, j = 1, \ldots, r$

$$
\begin{aligned}
Y_{tj} &= \boldsymbol{F}_t'\boldsymbol{\theta}_{tj} + \nu_{tj}, & \nu_{tj} &\sim \mathrm{N}[0, V_t\sigma_j^2],\\
\boldsymbol{\theta}_{tj} &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1,j} + \omega_{tj}, & \omega_{tj} &\sim \mathrm{N}[0, \boldsymbol{W}_t\sigma_j^2].
\end{aligned}
$$

The defining components $\boldsymbol{F}_t$, $\boldsymbol{G}_t$, V_t, and $\boldsymbol{W}_t$ are all the same (common) for each of the r series. With the following

$$
\begin{aligned}
\boldsymbol{Y}_t &= (Y_{t1}, \ldots, Y_{tr})', & \dim(\boldsymbol{Y}_t) &= r \times 1,\\
\boldsymbol{\nu}_t &= (\nu_{t1}, \ldots, \nu_{tr})', & \dim(\boldsymbol{\nu}_t) &= r \times 1,\\
\Theta_t &= (\boldsymbol{\theta}_{t1}, \ldots, \boldsymbol{\theta}_{tr}), & \dim(\Theta_t) &= n \times r,\\
\boldsymbol{\Omega}_t &= (\omega_{t1}, \ldots, \omega_{tr}), & \dim(\boldsymbol{\Omega}_t) &= n \times r,
\end{aligned}
$$

the Common Components Model is defined as

$$\boldsymbol{Y}'_t = \boldsymbol{F}'_t\Theta_t + \boldsymbol{\nu}'_t, \qquad \boldsymbol{\nu}_t \sim \mathrm{N}[0, V_t\Sigma], \tag{2.7}$$

$$\Theta_t = \boldsymbol{G}_t\Theta_{t-1} + \boldsymbol{\Omega}_t, \qquad \boldsymbol{\Omega}_t \sim \mathrm{N}[\mathbf{O}, \boldsymbol{W}_t, \Sigma], \tag{2.7'}$$

$$(\Theta_0, \Sigma | D_0) \sim \mathrm{NW}^{-1}_{n_0}[\boldsymbol{m}_0, \boldsymbol{C}_0, \boldsymbol{S}_0],$$

where $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$, $1 \le i, j \le r$, n_0, $\boldsymbol{m}_0$, $\boldsymbol{C}_0$, $\boldsymbol{S}_0$ are known quantities, and "$\mathrm{NW}^{-1}_{n_0}$" denotes the joint normal inverse Wishart distribution with n_0 degrees of freedom (see Section 4.2). V_t is assumed known and often it is set to 1. The following theorem is based on the matrix normal/inverse Wishart/matrix T distribution theory.

Theorem 2.4. *One-step forecast and posterior distributions in the CCM are given, for each t, as follows.*

(a) Posterior at $t-1$:

$$(\Theta_{t-1}, \Sigma | D_{t-1}) \sim NW^{-1}_{n_{t-1}}[\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}, \boldsymbol{S}_{t-1}],$$

for some known quantities $\boldsymbol{m}_{t-1}$, $\boldsymbol{C}_{t-1}$, $\boldsymbol{S}_{t-1}$, and n_{t-1}.

(b) Prior at t:

$$(\Theta_t, \Sigma | D_{t-1}) \sim NW^{-1}_{n_{t-1}}[\boldsymbol{a}_t, \boldsymbol{R}_t, \boldsymbol{S}_{t-1}],$$

where

$$\boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1} \qquad \text{and} \qquad \boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}'_t + \boldsymbol{W}_t.$$

(c) One-step forecast:

$$(\boldsymbol{Y}_t | \Sigma, D_{t-1}) \sim N[\boldsymbol{f}_t, Q_t\Sigma],$$

with marginal

$$(\boldsymbol{Y}_t|D_{t-1}) \sim T_{n_{t-1}}[\boldsymbol{f}_t, Q_t\boldsymbol{S}_{t-1}],$$

where

$$\boldsymbol{f}_t' = \boldsymbol{F}_t'\boldsymbol{a}_t \qquad \text{and} \qquad Q_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + V_t.$$

(d) Posterior at t:

$$(\Theta_t, \Sigma|D_t) \sim NW_{n_t}^{-1}[\boldsymbol{m}_t, \boldsymbol{C}_t, \boldsymbol{S}_t],$$

with

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t\boldsymbol{e}_t', \qquad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{A}_t'Q_t,$$

$$n_t = n_{t-1} + 1 \qquad \text{and} \qquad \boldsymbol{S}_t = n_t^{-1}[n_{t-1}\boldsymbol{S}_{t-1} + \boldsymbol{e}_t\boldsymbol{e}_t'/Q_t],$$

where

$$\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_t/Q_t \qquad \text{and} \qquad \boldsymbol{e}_t = \boldsymbol{Y}_t - \boldsymbol{f}_t.$$

Proof. See [34]. □

Identities

(1) $Q_t = (1 - \boldsymbol{F}_t'\boldsymbol{A}_t)^{-1}V_t$;

(2) $\boldsymbol{A}_t = \boldsymbol{C}_t\boldsymbol{F}_t/V_t$;

(3) $\boldsymbol{C}_t^{-1} = \boldsymbol{R}_t^{-1} + \boldsymbol{F}_t\boldsymbol{F}_t'/V_t$;

(4) $\boldsymbol{C}_t^{-1}\boldsymbol{m}_t = \boldsymbol{R}_t^{-1}\boldsymbol{a}_t + \boldsymbol{F}_t\boldsymbol{Y}_t'/V_t$.

Proof. (1) From the updatings of Q_t and $\boldsymbol{A}_t$ of the above theorem it is $(1 - \boldsymbol{F}_t'\boldsymbol{A}_t)Q_t = V_t$ and (1) follows. (2) first note

$$\boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{A}_t/V_t = \boldsymbol{A}_t\boldsymbol{A}_t'\boldsymbol{F}_tQ_t/V_t. \tag{2.8}$$

Then, the updatings of $\boldsymbol{C}_t$ and $\boldsymbol{A}_t$ of Theorem 2.4 imply

$$\boldsymbol{A}_t = \boldsymbol{C}_t\boldsymbol{F}_t/V_t - \boldsymbol{C}_t\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{A}_t/V_t + \boldsymbol{A}_t\boldsymbol{A}_t'\boldsymbol{F}_tQ_t/V_t - \boldsymbol{A}_t\boldsymbol{A}_t'\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{A}_tQ_t/V_t$$

and using equation (2.8) we have (2). (3) Using the updatings of $\boldsymbol{C}_t$ and $\boldsymbol{A}_t$ of Theorem 2.4 together with (2)

$$\begin{aligned}(\boldsymbol{R}_t^{-1} + \boldsymbol{F}_t\boldsymbol{F}_t'/V_t)\boldsymbol{C}_t &= \boldsymbol{I} - \boldsymbol{F}_t\boldsymbol{A}_t' + \boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{C}_t/V_t \\ &= \boldsymbol{I}.\end{aligned}$$

Similarly it can be proved that $\boldsymbol{C}_t(\boldsymbol{R}_t^{-1} + \boldsymbol{F}_t\boldsymbol{F}_t'/S_{t-1}) = \boldsymbol{I}$, and the result follows. (4) From (2)

$$\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{A}_t/V_t + \boldsymbol{R}_t^{-1}\boldsymbol{A}_t = \boldsymbol{F}_t/V_t. \tag{2.9}$$

Now using the updatings of $\boldsymbol{m}_t$, $\boldsymbol{C}_t$, $\boldsymbol{f}_t$ of Theorem 2.4 together with (3)

$$\boldsymbol{C}_t^{-1}\boldsymbol{m}_t = \boldsymbol{R}_t^{-1}\boldsymbol{a}_t + \boldsymbol{R}_t^{-1}\boldsymbol{A}_t\boldsymbol{e}_t' + \boldsymbol{F}_t\boldsymbol{f}_t'/V_t + \boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{A}_t\boldsymbol{e}_t'/V_t,$$

which with (2.9) establishes (4). □

The CCM is a special case of the general multivariate model. To see this, consider the CCM as defined by equations (2.7), (2.7′) and apply the following transormation

$$\begin{aligned}&\boldsymbol{Y}_t^* = \mathrm{vec}(\boldsymbol{Y}_t'), &&\boldsymbol{F}_t^* = \boldsymbol{I}_r \otimes \boldsymbol{F}_t, &&\boldsymbol{\theta}_t^* = \mathrm{vec}(\boldsymbol{\Theta}_t),\\ &\boldsymbol{\nu}_t^* = \mathrm{vec}(\boldsymbol{\nu}_t), &&\boldsymbol{G}_t^* = \boldsymbol{I}_r \otimes \boldsymbol{G}_t, &&\boldsymbol{\omega}_t^* = \mathrm{vec}(\boldsymbol{\Omega}_t),\end{aligned}$$

where "vec", "⊗" denote the vec-operator and the Kronecker product of matrices respectively (see Appendix A.2). Under this transformation the CCM can be written as

$$\begin{aligned}&\boldsymbol{Y}_t^* = \boldsymbol{F}_t^{*'}\boldsymbol{\theta}_t^* + \boldsymbol{\nu}_t^*, &&\boldsymbol{\nu}_t^* \sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{\Sigma} \otimes V_t],\\ &\boldsymbol{\theta}_t^* = \boldsymbol{G}_t^*\boldsymbol{\theta}_{t-1}^* + \boldsymbol{\omega}_t^*, &&\boldsymbol{\omega}_t^* \sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{W}_t],\end{aligned}$$

which is essentially embedded to model (2.5), (2.5′).

According to [34], the above methodology can be generalised to the case when $\boldsymbol{Y}_t$ is an $r \times m$ matrix. Then, all the above analysis remains, with $\boldsymbol{V}_t$ an $m \times m$ variance matrix and $\boldsymbol{Q}_t$ an $m \times m$ matrix. In this case identities (2) - (4) do not hold, since in general $\boldsymbol{F}_t'\boldsymbol{A}_t \neq \boldsymbol{A}_t'\boldsymbol{F}_t$.

These models are very limited due to the common components structure. The r univariate series must be similar, and modelling several univariate series independently to each other seems more sensible. Barbosa and Harrison ([4]) criticised this model and proposed an approximation to the general multivariate DLM. A brief discussion and extensions of this model appear in Chapter 4.

2.4 Specification of the Evolution Variance

This is a brief discussion of the discounting approach that first appeared in [1, 2] and further explained in [51, chapter 6].

2.4.1 Single Discounting

Consider the general multivariate model (model (2.5), (2.5′)) where $\boldsymbol{V}_t$ is known. In the context of Theorem 2.3, define $\boldsymbol{P}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t'$ to be the variance of the quantity $\boldsymbol{G}_t\boldsymbol{\theta}_{t-1}$ given D_{t-1}. For a static model $\boldsymbol{W}_t$ is zero, so that $\boldsymbol{R}_t = \boldsymbol{P}_t$. Or in precision terms $\boldsymbol{R}_t^{-1} = \boldsymbol{P}_t^{-1}$. However, if the model is only locally appropriate due to the non-zero variance $\boldsymbol{W}_t$, the precision $\boldsymbol{R}_t^{-1}$ is reduced relative to $\boldsymbol{P}_t^{-1}$. The easiest and most natural way to express this relation is by

$$\boldsymbol{R}_t^{-1} = \delta \boldsymbol{P}_t^{-1},$$

where δ is a discount factor $0 < \delta \leq 1$. This implies

$$\boldsymbol{R}_t = \frac{1}{\delta} \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t'$$

and from $\boldsymbol{R}_t = \boldsymbol{P}_t + \boldsymbol{W}_t$, the variance $\boldsymbol{W}_t$ is specified as

$$\boldsymbol{W}_t = \frac{1-\delta}{\delta} \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t.$$

Notice that if $\delta = 1$, then $\boldsymbol{W}_t = \mathbf{O}$ and we have a static model. If, on the the other extreme, $\delta \approx 0$ the elements of $\boldsymbol{W}_t$ will be very large, returning a very unstable model. In the first case the model is said to be very durable whereas in the second case it is not durable and probably useless. This method of discounting is referred to as **single discounting** because a single discount factor has been considered. The choice of discount factor and its relationship with intervention is discussed in [43].

2.4.2 Component Discounting

Single discounting is usually not appropriate, especially when the model comprises different components that have different durability. This was shown in [15]. For example, usually seasonal and trend components are conditionally independent and require different discount factors.

Consider m multivariate DLMs, $\mathcal{M}_i$ $(i = 1, \ldots, m)$, with $n_i \times 1$ state vectors $\boldsymbol{\theta}_{it}$, and evolution errors $\boldsymbol{\omega}_{it}$. Write for all $i = 1, \ldots, m$,

$$\mathcal{M}_i : \qquad \{\boldsymbol{F}_{it}, \boldsymbol{G}_{it}, \boldsymbol{V}_{it}, \boldsymbol{W}_{it}\},$$

where the matrices $\boldsymbol{F}_{it}$, $\boldsymbol{G}_{it}$, $\boldsymbol{V}_{it}$ are known for all i, t, and the $n_i \times n_i$ matrices $\boldsymbol{W}_{it}$ are specified via a single discount factor. So

$$\boldsymbol{W}_{it} = \frac{1-\delta_i}{\delta_i}\boldsymbol{G}_{it}\boldsymbol{C}_{i,t-1}\boldsymbol{G}'_{it},$$

where δ_i, $(i = 1\ldots,m)$, are any m discount factors. Then, a component DLM can be defined by the quadruples

$$\{\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_t, \boldsymbol{W}_t\}$$

and $\boldsymbol{\theta}_t$ as its $n \times 1$ state vector, such that $n = \sum_{i=1}^{m} n_i$, where

$$\boldsymbol{F}'_t = (\boldsymbol{F}'_{1t}, \ldots, \boldsymbol{F}'_{mt}), \qquad \boldsymbol{G}_t = \text{block diag}\{\boldsymbol{G}_{1t}, \ldots, \boldsymbol{G}_{mt}\},$$

$$\boldsymbol{\theta}'_t = (\boldsymbol{\theta}'_{1t}, \ldots, \boldsymbol{\theta}'_{mt}), \qquad \boldsymbol{W}_t = \text{block diag}\{\boldsymbol{W}_{1t}, \ldots, \boldsymbol{W}_{mt}\}.$$

This DLM is called the superposition of the models $\mathcal{M}_1, \ldots \mathcal{M}_m$ (see the following section). A discussion of the general approach of superposition is in Section 2.6. This method of discounting is referred to as the **component discounting** because of the structure of the component model. If $m = 2$ it is also used the term of **double discounting.** Any DLM, with evolution matrix specified via the discounting approach, is called a discount DLM. It is also worth noting that although the above concept is built upon the known variance DLM, it generalises to unknown variances.

2.5 Superposition and Decomposition of DLMs

The principles of **superposition** and **decomposition** of a DLM are complementary. The former refers to the construction of a complex DLM from simple component DLMs and the later refers to the identification of components of a given DLM. Let $\boldsymbol{f}_t(k)$ be the k-step forecast function of model

(2.5), (2.5′), given by $\boldsymbol{f}_t(k) = \mathrm{E}[\boldsymbol{Y}_{t+k}|D_t]$. The next theorem gives the main result, which is of key importance.

Theorem 2.5. *Consider m multivariate $r \times 1$ times series $\boldsymbol{Y}_{it}$ generated by DLMs*

$$\mathcal{M}_i: \quad \{\boldsymbol{F}_{it}, \boldsymbol{G}_{it}, \boldsymbol{V}_{it}, \boldsymbol{W}_{it}\},$$

for $i = 1, \ldots, m$. In $\mathcal{M}_i$, the state vector $\boldsymbol{\theta}_{it}$ is of dimension n_i, and the observation and evolution errors are respectively $\boldsymbol{\nu}_{it}$ and $\boldsymbol{\omega}_{it}$. The state vectors $\boldsymbol{\theta}_{1t}, \ldots, \boldsymbol{\theta}_{mt}$ are distinct.

(i) If for all distinct $i \neq j$, the series $\boldsymbol{\nu}_{it}$ and $\boldsymbol{\omega}_{it}$ are mutually independent of the series $\boldsymbol{\nu}_{jt}$ and $\boldsymbol{\omega}_{jt}$, then the $r \times 1$ series

$$\boldsymbol{Y}_t = \sum_{i=1}^{m} \boldsymbol{Y}_{it}$$

has the following DLM

$$\mathcal{M}: \quad \{\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_t, \boldsymbol{W}_t\},$$

where $n = \sum_{i=1}^{m} n_i$ and the $n \times 1$ state vector $\boldsymbol{\theta}_t$ and quadruple are given by

$$\boldsymbol{\theta}_t' = (\boldsymbol{\theta}_{1t}', \ldots, \boldsymbol{\theta}_{mt}'), \qquad \boldsymbol{F}_t' = (\boldsymbol{F}_{1t}', \ldots, \boldsymbol{F}_{mt}'),$$

$$\boldsymbol{G}_t = \textit{block diag}\{\boldsymbol{G}_{1t}, \ldots, \boldsymbol{G}_{mt}\}, \qquad \boldsymbol{W}_t = \textit{block diag}\{\boldsymbol{W}_{1t}, \ldots, \boldsymbol{W}_{mt}\},$$

and

$$\boldsymbol{V}_t = \sum_{i=1}^{m} \boldsymbol{V}_{it}.$$

(ii) If in addition, the series $\boldsymbol{\nu}_{it}$ and $\boldsymbol{\omega}_{it}$ are internally independent and each have a multivariate joint normal distribution, for all $i = 1, \ldots, m$, then

the forecast function of $\boldsymbol{Y}_t$, $\boldsymbol{f}_t(k)$, is given by

$$\boldsymbol{f}_t(k) = \sum_{i=1}^{m} \boldsymbol{f}_{it}(k),$$

where $\boldsymbol{f}_{it}(k)$ denotes the forecast function of the model $\mathcal{M}_i$.

Proof. The proof is completely analogous to the univariate case given in [51, chapter 6], the only difference being that the quantities ν_{it}, Y_t, Y_{it}, V_t, V_{it}, $f_t(k)$, and $f_{it}(k)$ of the above reference are replaced by the multivariate quantities $\boldsymbol{\nu}_{it}$, $\boldsymbol{Y}_t$, $\boldsymbol{Y}_{it}$, $\boldsymbol{V}_t$, $\boldsymbol{V}_{it}$, $\boldsymbol{f}_t(k)$, and $\boldsymbol{f}_{it}(k)$. $\square$

Model $\mathcal{M}$ is said to be the superposition of m DLMs, $\mathcal{M}_i$, while $\mathcal{M}_i$, $(i = 1, \ldots, m)$, the decomposition of $\mathcal{M}$.

The above theorem provides key results for the model building. If a complex model is required we divide it into simpler components, we build these component models and then we formulate the general model. If a special component of a complex model is of interest we work with this without being bothered with the whole model. Examples of these methods are explicitly explored in [51, chapters 6,7,8].

2.6 Observability and Model Design

Model design and parameter economy are fundamental to model building. Chatfield, as a discussant in [19], criticised dynamic models due to the difficulty in choosing the appropriate parameter elements $\boldsymbol{F}_t$, $\boldsymbol{G}_t$. [51] gives a complete account of model design, for the univariate case. The interested reader is referred to Chapters 5, 6 for a general treatment, to Chapters 7, 8 for the important classes of Polynomial Trend and Seasonal Models, and to

Chapters 9, 10 for Regression and special topics on univariate DLMs. This section aims to providing the multivariate analogue of the relevant model design developed in Chapters 5, 6 of the above reference.

2.6.1 Observability

Consider any multivariate DLM $\{\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_t, \boldsymbol{W}_t\}$, where the $r \times r$ matrix $\boldsymbol{V}_t$ is either known or unknown and $\boldsymbol{W}_t$ is appropriately specified, either in advance or using the discounting approach of Section 2.4. Consider first the case of $\boldsymbol{W}_t = \mathbf{O}$. Define the mean response vector $\boldsymbol{\mu}_t = \mathrm{E}[\boldsymbol{Y}_{t+k}|\boldsymbol{\theta}_{t+k}] = \boldsymbol{F}'_{t+k}\boldsymbol{G}_{t+k}\boldsymbol{G}_{t+k-1}\cdots\boldsymbol{G}_{t+1}\boldsymbol{\theta}_t$ and $\boldsymbol{\mu}_t^{*\prime} = (\boldsymbol{\mu}'_t, \boldsymbol{\mu}'_{t+1}, \ldots, \boldsymbol{\mu}'_{t+n-1})$. Then parameter economy suggests that the mapping $\boldsymbol{\mu}_t^* = \boldsymbol{T}_t\boldsymbol{\theta}_t$ is a bijection, where $\boldsymbol{T}_t$ is the $(nr) \times n$ observability matrix

$$\boldsymbol{T}_t = \begin{pmatrix} \boldsymbol{F}'_t \\ \boldsymbol{F}'_{t+1}\boldsymbol{G}_{t+1} \\ \vdots \\ \boldsymbol{F}'_{t+n-1}\prod_{i=1}^{n-1}\boldsymbol{G}_{t+n-i} \end{pmatrix}. \tag{2.10}$$

The quantities $\boldsymbol{\mu}_{t+i}$ $(i = 0, \ldots, n-1)$ are linearly independent if and only if $\boldsymbol{T}_t$ is of full rank. Otherwise the parametrisation can be reduced to the first m independent elements of $\boldsymbol{\mu}_{t+i}$ $(i = 0, \ldots, m-1)$ such that $\mathrm{rank}(\boldsymbol{T}_t) = m$.

The above approach of parametric parsimony motivates the case of $\boldsymbol{W}_t > \mathrm{O}$.

Definition 2.1. *Any DLM* $\{\boldsymbol{F}_t, \boldsymbol{G}_t, \cdot, \cdot\}$ *is* **observable** *if and only if the* $(nr) \times n$ **observability matrix** $\boldsymbol{T}_t$ *in (2.10) has full rank* n *for every* t.

Examples

(1) Consider the DLM $\{\boldsymbol{F}_t, \boldsymbol{G}_t, \cdot, \cdot\}$, where $r = n$ and $|\boldsymbol{F}_t| \neq 0$. Then, the model is always observable, since $\text{rank}(\boldsymbol{T}_t) = n$, for all t. The univariate analogue of this is the DLM $\{F_t, G_t, \cdot, \cdot\}$, where $r = n = 1$ and $T_t = F_t \neq 0$.

(2) Consider the univariate deterministic DLM

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0.8 \end{pmatrix}, 0, \mathbf{O} \right\}. \tag{2.11}$$

The observability matrix $\boldsymbol{T}$,

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

has rank 1 and model (2.11) can be reduced to the deterministic

$$\{1, 1, 0, 0\}.$$

Now consider the alternative DLM

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0.8 \end{pmatrix}, 1, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}, \tag{2.12}$$

with the same observability matrix as model (2.11). Write this model as

$$Y_t = \theta_t + \nu_t, \qquad \nu_t \sim \text{N}[0, 1], \tag{2.13}$$

$$\theta_t = \theta_{t-1} + \omega_t, \qquad \omega_t \sim \text{N}[0, 1], \tag{2.13$'$}$$

$$\psi_t = 0.8\psi_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{N}[0, 1], \tag{2.13$''$}$$

where $\boldsymbol{\theta}_t = (\theta_t, \psi_t)'$ and $\boldsymbol{\omega}_t = (\omega_t, \epsilon_t)'$. The model can be reduced to the first-order constant DLM

$$\{1, 1, 1, 1\},$$

with the single parameter θ_t, if forecasting the series Y_t is the goal. But if there is interest in ψ_t, writing equation (2.13″) recursively and taking its variance

$$V[\psi_t|D_t] = \sum_{i=0}^{t-1} 0.64^i + 0.64^t V[\psi_0|D_0],$$

which, with $(\psi_0|D_0) \sim N[a_0, c_0]$, $(C_0 < \infty)$, provide the limiting distribution of $(\psi_t|D_t)$, as

$$(\psi_t|D_t) \sim \text{Asymptotically } N[0, 2.77].$$

The last example shows that the modeller has to set up the targets. The lack of observability does not always mean that a reduction in parameterization is needed. However, when forecasting is considered, it is recommended that an observable model is used.

Let $\mathcal{B}(x, \epsilon)$ denote the open neighborhood of x, for any $\epsilon > 0$ and x be any point of the real line.

Definition 2.2. *Any DLM $\{\boldsymbol{F}_t, \boldsymbol{G}_t, \cdot, \cdot\}$ is said to be* **locally observable** *on a neighborhood of t_0 if and only if there exists $\epsilon > 0$ such that the DLM $\{\boldsymbol{F}_t, \boldsymbol{G}_t, \cdot, \cdot\}$ is observable for all $t \in \mathcal{B}(t_0, \epsilon)$.*

Modelling $\boldsymbol{F}_t = \boldsymbol{F}$ and $\boldsymbol{G}_t = \boldsymbol{G}$ results in many popular "time series" models. It can be shown that any ARIMA model may be written as constant DLM $\{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{V}, \boldsymbol{W}\}$. The correspondent ARIMA predictor corresponds to the limiting form of that of the DLM, see [51, chapter 5].

Definition 2.3. *The general multivariate DLM $\{\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_t, \boldsymbol{W}_t\}$ is referred to as a* **time series DLM** *or* **TSDLM** *if and only if the pair $\{\boldsymbol{F}_t, \boldsymbol{G}_t\}$ is constant over time. If in addition the variances $\boldsymbol{V}_t$, $\boldsymbol{W}_t$ are constant, then it is referred to as* **constant DLM**.

It is easy to verify that for any TSDLM $\{\boldsymbol{F}, \boldsymbol{G}, \cdot, \cdot\}$, it is $\boldsymbol{f}_t(k) = \boldsymbol{F}'\boldsymbol{G}^k\boldsymbol{m}_t$, with $\boldsymbol{m}_t$ as in Theorem 2.3 (page 17).

Our attention focuses on multivariate TSDLMs, and it may be noted for that for a multivariate TSDLM $\{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{V}_t, \boldsymbol{W}_t\}$, a necessary and sufficient condition for observability is that the univariate TSDLM $\{\boldsymbol{F}\boldsymbol{l}, \boldsymbol{G}, \boldsymbol{l}'\boldsymbol{V}_t\boldsymbol{l}, \boldsymbol{W}_t\}$ is observable for every non-zero $r \times 1$ vector $\boldsymbol{l}$.

2.6.2 Similar and Equivalent Models

Similar TSDLMs have similar evolution matrices, see Appendix A.1.

Consider two observable TSDLMs, with quadruples

$$\begin{aligned} \mathcal{M}: &\quad \{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{V}_t, \boldsymbol{W}_t\}, \\ \mathcal{M}_1: &\quad \{\boldsymbol{F}_1, \boldsymbol{G}_1, \boldsymbol{V}_{1t}, \boldsymbol{W}_{1t}\}. \end{aligned}$$

Denote the forecast functions of $\mathcal{M}$, $\mathcal{M}_1$ as $\boldsymbol{f}_t(k)$, $\boldsymbol{f}_{1t}(k)$ respectively. Then we have the following definition.

Definition 2.4. *$\mathcal{M}$ and $\mathcal{M}_1$ are said to be similar if and only if the evolution matrices $\boldsymbol{G}$, $\boldsymbol{G}_1$ are similar. The notation is $\mathcal{M} \sim \mathcal{M}_1$.*

A direct consequence of the definition is that the matrices $\boldsymbol{G}$ and $\boldsymbol{G}_1$ have identical eigenvalues. Suppose that they have n distinct eigenvalues namely $\lambda_1, \dots, \lambda_n$, so that both $\boldsymbol{G}$, $\boldsymbol{G}_1$ are diagonalisable. Define the diagonal matrix Λ as $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\}$. Then, there exists non-singular $n \times n$ matrix $\boldsymbol{E}$ such that $\boldsymbol{G} = \boldsymbol{E}\Lambda\boldsymbol{E}^{-1}$ and $\boldsymbol{G}^k = \boldsymbol{E}\Lambda^k\boldsymbol{E}^{-1}$, for all k. So at time t, given D_t, the forecast function of $\mathcal{M}$ is

$$\boldsymbol{f}_t(k) = \sum_{i=1}^{n} \lambda_i^k \boldsymbol{a}_{ti},$$

where $\boldsymbol{a}_{ti}$ $(i = 1, \ldots, n)$ are functions of the known information D_t, and do not depend on k and λ_j, $(j = 1, \ldots, n)$.

Since $\boldsymbol{G}_1$ has eigenvalues $\lambda_1, \ldots, \lambda_n$, there exists a non-singular $n \times n$ matrix $\boldsymbol{H}$ not involving any of λ_j, such that $\boldsymbol{G}_1 = \boldsymbol{H}\boldsymbol{G}\boldsymbol{H}^{-1} = \boldsymbol{H}\boldsymbol{E}\Lambda\boldsymbol{E}^{-1}\boldsymbol{H}^{-1}$ and the forecast function of $\mathcal{M}_1$ will be

$$\boldsymbol{f}_{1t}(k) = \sum_{i=1}^{n} \lambda_i^k \boldsymbol{b}_{ti},$$

where $\boldsymbol{b}_{ti}$, $(i = 1, \ldots, n)$, does not depend on k and λ_j, $(j = 1, \ldots, n)$.

The conclusion is that similarity implies that the forecast functions $\boldsymbol{f}_t(k)$, $\boldsymbol{f}_{1t}(k)$ have the same algebraic form. This is true in general, even if the eigenvalues are not all distinct. Note that if $\boldsymbol{G}$ is block diagonal, $\boldsymbol{G}_1$ is block diagonal with similar block elements to $\boldsymbol{G}$. For more details about similarity see ([51, chapter 5]). Appendix A.1 provides the necessary preliminaries of matrix algebra, but for a thorough treatment the reader is referred to [24, chapter 21].

Consider the model $\mathcal{M}_1$ defined by equations

$$\begin{aligned} \boldsymbol{Y}_t &= \boldsymbol{F}_1'\boldsymbol{\theta}_{1t} + \boldsymbol{\nu}_{1t}, & \boldsymbol{\nu}_{1t} &\sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{V}_{1t}], \\ \boldsymbol{\theta}_{1t} &= \boldsymbol{G}_1\boldsymbol{\theta}_{1,t-1} + \boldsymbol{\omega}_{1t}, & \boldsymbol{\omega}_{1t} &\sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{W}_{1t}], \end{aligned}$$

and $\boldsymbol{H}$ a non-singular $n \times n$ matrix. Then $\mathcal{M}_1$ can be transformed to another model $\mathcal{M}$ via the linear transformation

$$\boldsymbol{\theta}_t = \boldsymbol{H}\boldsymbol{\theta}_{1t},$$

where $\boldsymbol{\theta}_t$ is the $n \times 1$ state vector of $\mathcal{M}$. Then, $\mathcal{M}$ is defined by

$$\begin{aligned} \boldsymbol{Y}_t &= \boldsymbol{F}'\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, & \boldsymbol{\nu}_t &\sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{V}_t], \\ \boldsymbol{\theta}_t &= \boldsymbol{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t &\sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{W}_t], \end{aligned}$$

where $\boldsymbol{F} = \boldsymbol{H}'^{-1}\boldsymbol{F}_1$, $\nu_t = \nu_{1t}$, $\boldsymbol{V}_t = \boldsymbol{V}_{1t}$, $\boldsymbol{G} = \boldsymbol{H}\boldsymbol{G}_1\boldsymbol{H}^{-1}$, $\boldsymbol{\omega}_t = \boldsymbol{H}\boldsymbol{\omega}_{1t}$, and $\boldsymbol{W}_t = \boldsymbol{H}\boldsymbol{W}_{1t}\boldsymbol{H}'$. It is also $\boldsymbol{m}_t = \boldsymbol{H}\boldsymbol{m}_{1t}$ and $\boldsymbol{G}^k = \boldsymbol{H}\boldsymbol{G}_1^k\boldsymbol{H}^{-1}$. The forecast functions are identical since

$$\boldsymbol{f}_{1t}(k) = \boldsymbol{F}_1'\boldsymbol{G}_1^k\boldsymbol{m}_{1t} = \boldsymbol{F}_1'\boldsymbol{H}^{-1}\boldsymbol{H}\boldsymbol{G}_1^k\boldsymbol{H}^{-1}\boldsymbol{H}\boldsymbol{m}_{1t} = \boldsymbol{F}'\boldsymbol{G}^k\boldsymbol{m}_t = \boldsymbol{f}_t(k). \quad (2.14)$$

This implies that if the observability matrices of $\mathcal{M}_1$, $\mathcal{M}$ are denoted by $\boldsymbol{T}_1$, $\boldsymbol{T}$ respectively, then it is

$$\boldsymbol{T}_1 = \boldsymbol{T}\boldsymbol{H}. \quad (2.15)$$

All the above are summarised in the following theorem.

Theorem 2.6. *Let $\boldsymbol{H}$ be a non-singular matrix, such that for the observable TSDLMs $\mathcal{M} = \{\boldsymbol{F}, \boldsymbol{G}, \cdot, \cdot\}$ and $\mathcal{M}_1 = \{\boldsymbol{F}_1, \boldsymbol{G}_1, \cdot, \cdot\}$, $\boldsymbol{F} = \boldsymbol{H}'^{-1}\boldsymbol{F}_1$ and $\boldsymbol{G} = \boldsymbol{H}\boldsymbol{G}_1\boldsymbol{H}^{-1}$. If $\boldsymbol{T}$, $\boldsymbol{T}_1$ are the respective observability matrices of $\mathcal{M}$ and $\mathcal{M}_1$, then*

$$\text{(i)}\quad \mathcal{M} \sim \mathcal{M}_1 \qquad \text{and} \qquad \text{(ii)}\quad \boldsymbol{H} = (\boldsymbol{T}'\boldsymbol{T})^{-1}\boldsymbol{T}'\boldsymbol{T}_1.$$

Proof. (i) Equation (2.14) guarantees similarity.
(ii) Observability implies that rank$(\boldsymbol{T}) = n$, so that the $n \times n$ matrix $\boldsymbol{T}'\boldsymbol{T}$ is non-singular. The proof follows on premultiplying equation (2.15) by $\boldsymbol{T}'$. □

Matrix $\boldsymbol{H}$ is called the similarity matrix. Now we are ready to define equivalence.

Definition 2.5. *Consider two similar observable TSDLMs $\mathcal{M}$ and $\mathcal{M}_1$ with similarity matrix $\boldsymbol{H} = (\boldsymbol{T}'\boldsymbol{T})^{-1}\boldsymbol{T}'\boldsymbol{T}_1$. Suppose $\mathcal{M} = \{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{V}_t, \boldsymbol{W}_t\}$ with initial moments $(\boldsymbol{m}_0, \boldsymbol{C}_0)$ and $\mathcal{M}_1 = \{\boldsymbol{F}_1, \boldsymbol{G}_1, \boldsymbol{V}_{1t}, \boldsymbol{W}_{1t}\}$ with initial moments $(\boldsymbol{m}_{10}, \boldsymbol{C}_{10})$. Then $\mathcal{M}$ and $\mathcal{M}_1$ are said to be* **equivalent**, *denoted by*

$\mathcal{M} \equiv \mathcal{M}_1$, *if and only if*

$$\boldsymbol{V}_t = \boldsymbol{V}_{1t} \qquad \textit{and} \qquad \boldsymbol{W}_t = \boldsymbol{H}\boldsymbol{W}_{1t}\boldsymbol{H}'$$

for all t*, with*

$$\boldsymbol{m}_0 = \boldsymbol{H}\boldsymbol{m}_{10} \qquad \textit{and} \qquad \boldsymbol{C}_0 = \boldsymbol{H}\boldsymbol{C}_{10}\boldsymbol{H}'.$$

Notice that the above definition is in accord with the univariate case, where $\boldsymbol{H} = \boldsymbol{T}^{-1}\boldsymbol{T}_1$. Equivalence for the multivariate case can be handled working always with the univariate model $\{\boldsymbol{F}\boldsymbol{l}, \boldsymbol{G}, \boldsymbol{l}'\boldsymbol{V}_t\boldsymbol{l}, \boldsymbol{W}_t\}$, instead of the $\{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{V}_t, \boldsymbol{W}_t\}$. However, here the direct multivariate approach is preferred for reasons that will be clear in the next section.

2.6.3 Specification of $\boldsymbol{F}$ and $\boldsymbol{G}$

The idea of model design is based on the superposition and decomposition principle, described in Section 2.5. Complicated models are decomposed into a set of simpler models known as **canonical** models. This section explores the canonical multivariate DLMs.

Definition 2.6. *The* $n \times n$ **Jordan block** *is defined, for real or complex* λ *as the* $n \times n$ *upper diagonal matrix*

$$\boldsymbol{J}_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & 0 & \cdots & 0 \\ 0 & 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}.$$

Suppose that G has one eigenvalue of multiplicity $n > 1$. For example it can be $G = \lambda I$. The following result states that if the model is observable G is necessarily similar to $J_n(\lambda)$.

Theorem 2.7. *If G has one eigenvalue, λ, of multiplicity n but is not similar to $J_n(\lambda)$, then any TSDLM $\{F, G, \cdot, \cdot\}$ is unobservable.*

Proof. Since G is not similar to the Jordan block, it will be similar to a Jordan form

$$J_s = \text{block diag}\{J_{n_1}(\lambda), \dots, J_{n_s}(\lambda)\},$$

for some $s \geq 2$, $\sum_{p=1}^{s} n_p = n$ and $n_p \geq 1$, $(p = 1, \dots, s)$. For each $1 \leq i \leq r$ let f_{ij} be any $n_p \times 1$ vector, and define the $n \times r$ design matrix F_s via $F_s' = (f_1', \dots, f_s')$, where the $n_p \times r$ matrices f_p, $(p = 1, \dots, s)$, are defined as $f_p = (f_{1n_p}, \dots, f_{rn_p})$. Then, the $i + k$ row of the observability matrix T of any model with design matrix F_s and system matrix J_s is

$$t_{i+k}' = (f_{in_1}' J_{n_1}^k(\lambda), \dots, f_{in_s}' J_{n_s}^k(\lambda)), \qquad (i = 1, \dots, r; k = 0, 1, \dots, n-1).$$

Define $m = \max\{n_1, \dots, n_s\}$, so that $m \leq n - s + 1$. Then, for any $m < k \leq n - 1$ it is

$$\sum_{j=0}^{k} \binom{k}{j} (-\lambda)^j t_{i+k-j}' = (f_{in_1}' J_{n_1}^k(0), \dots, f_{in_s}' J_{n_s}^k(0))$$

and using Appendix A.1

$$\sum_{j=0}^{k} \binom{k}{j} (-\lambda)^j t_{i+k-j}' = 0',$$

for $i = 1, \dots, r$ and $n_p \leq m < k$, $(p = 1, \dots, s)$.

This implies that T is of less than full rank, having at most $m < n$ linearly independent rows, and so the DLM is not observable. □

Corollary 2.1. *If $\boldsymbol{G}$ and $\boldsymbol{G}_1$ each have a single eigenvalue λ of multiplicity n and for some $\boldsymbol{F}$ and $\boldsymbol{F}_1$, the models $\{\boldsymbol{F},\boldsymbol{G},\cdot,\cdot\}$ and $\{\boldsymbol{F}_1,\boldsymbol{G}_1,\cdot,\cdot\}$ are observable, then $\boldsymbol{G}$ is similar to $\boldsymbol{G}_1$ and the models are similar.*

The next theorem identifies the simplest class of observable DLMs.

Theorem 2.8. *Any TSDLM $\{\boldsymbol{F},\boldsymbol{J}_n(\lambda),\cdot,\cdot\}$ is observable if and only if there is at least one non-zero element in the first row of $\boldsymbol{F}$.*

Proof. Let $\boldsymbol{F} = (\boldsymbol{f}_1,\ldots,\boldsymbol{f}_r) = \{f_{ij}\}$ $(i = 1,\ldots,n; j = 1,\ldots,r)$, where $\boldsymbol{f}_j = (f_{1j},\ldots,f_{nj})'$ are $n \times 1$ vectors. Let $\boldsymbol{A}$ be an $(nr) \times n$ matrix, where $\boldsymbol{A}' = (\boldsymbol{A}_1',\ldots,\boldsymbol{A}_r')$, and the $n \times n$ matrices $\boldsymbol{A}_j$ $(j = 1,\ldots,r)$ have rows

$$\boldsymbol{a}'_{j;k+1} = \sum_{m=0}^{k} \binom{k}{m} (-\lambda)^{k-m} \boldsymbol{t}'_{j+m}, \qquad (k = 0,1,\ldots,n-1),$$

where $\boldsymbol{t}'_{j+m}$ $(j = 1,\ldots,r; m = 0,1,\ldots,n-1)$ are the rows of the matrix $\boldsymbol{F}'\boldsymbol{J}_n^m(\lambda)$. Then

$$\begin{aligned} \boldsymbol{a}'_{j;k+1} &= \boldsymbol{f}'_j[\boldsymbol{J}_n(\lambda) - \lambda \boldsymbol{I}_n]^k \\ &= (0,\ldots,0,f_{1j},\ldots,f_{n-k,j}), \end{aligned}$$

having k leading zeros. Thus $\boldsymbol{A}_j$ is an upper triangular matrix with leading diagonal $(f_{1j},\ldots,f_{1j})'$ and with determinant f_{1j}^n. So $\operatorname{rank}(\boldsymbol{A}) = n$ if and only if $\exists j \in \mathbb{N}$, $(1 \leq j \leq r)$, such that $f_{1j} \neq 0$. The rows of $\boldsymbol{A}$ are constructed as linearly independent linear combinations of those of $\boldsymbol{T}$. So $\boldsymbol{A}$ and $\boldsymbol{T}$ have the same rank and the result follows. $\square$

The simplest class of observable TSDLMs, having one eigenvalue, λ, of multiplicity n consists of TSDLMs $\{\boldsymbol{E}_{n,i}, \boldsymbol{J}_n(\lambda), \cdot, \cdot\}$, where

$$\boldsymbol{E}_{n,i} = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

where the "1" is located in the ith column ($i = 1, \ldots, r$).

Suppose now that $\boldsymbol{G}$ has s real distinct eigenvalues $\lambda_1, \ldots, \lambda_s$ with λ_i having multiplicity $r_i \geq 1$, so that $\sum_{i=1}^{s} r_i = n$. From Appendix A.1, it follows that $\boldsymbol{G}$ is similar to the block diagonal Jordan form matrix

$$\boldsymbol{J} = \text{block diag}\{\boldsymbol{J}_{r_1}(\lambda_1), \ldots, \boldsymbol{J}_{r_s}(\lambda_s)\} \tag{2.16}$$

defined by the superposition of s Jordan blocks, one for each eigenvalue and having dimension given by the multiplicity of the eigenvalue.

Suppose that the evolution matrix $\boldsymbol{W}_t$ is modelled as

$$\boldsymbol{W}_t = \text{block diag}\{\boldsymbol{W}_{r_1,t}, \ldots, \boldsymbol{W}_{r_s,t}\}, \tag{2.17}$$

for some known $r_i \times r_i$, ($i = 1, \ldots, s$) variance matrices $\boldsymbol{W}_{r_i,t}$.

Theorem 2.9. *Any TSDLM $\{\boldsymbol{F}, \boldsymbol{J}, \cdot, \boldsymbol{W}_t\}$ with $\boldsymbol{J}, \boldsymbol{W}_t$ as defined in equations (2.16), (2.17), is observable if and only if there is at least one non-zero element in every $r_{i-1} + 1$ ($i = 1, \ldots, s; r_0 = 0$) row of $\boldsymbol{F}$.*

Proof. Let $\boldsymbol{F}' = (\boldsymbol{F}'_1, \ldots, \boldsymbol{F}'_s)$, where $\boldsymbol{F}_i$ ($i = 1, \ldots, s$) is an $r_i \times r$ matrix. Then, according to Section 2.5 the TSDLM $\{\boldsymbol{F}, \boldsymbol{J}, \cdot, \boldsymbol{W}_t\}$ can be decomposed into s TSDLMs $\{\boldsymbol{F}_i, \boldsymbol{J}_{r_i}(\lambda_i), \cdot, \boldsymbol{W}_{r_i,t}\}$ ($i = 1, \ldots, s$), each one of these satisfy the conditions of Theorem 2.8 and they are observable if and only if there is at least one non-zero element in the first row of $\boldsymbol{F}_i$. The proof is complete by noticing that the TSDLM $\{\boldsymbol{F}, \boldsymbol{J}, \cdot, \boldsymbol{W}_t\}$ is observable when all TSDLMs $\{\boldsymbol{F}_i, \boldsymbol{J}_{r_i}(\lambda_i), \cdot, \boldsymbol{W}_{r_i,t}\}$ are observable. □

The simplest class of observable TSDLMs, having s eigenvalues, consists of TSDLMs $\{\boldsymbol{E}_{nr}, \boldsymbol{J}, \cdot, \boldsymbol{W}_t\}$, where

$$\boldsymbol{E}'_{nr} = (\boldsymbol{E}'_{r_1,i_1}, \ldots, \boldsymbol{E}'_{r_s,i_s}), \tag{2.18}$$

where $1 \leq i_j \leq r_j$, $1 \leq j \leq s$.

The above development motivates the case of a general evolution matrix $\boldsymbol{W}_t$.

Similar analyses apply to the more general case where there may exist complex eigenvalues as well as real ones. These cases are not covered here and the interested reader is referred to [51, chapter 5].

Definition 2.7. *Let $\mathcal{M} = \{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{V}_t, \boldsymbol{W}_t\}$ be any observable TSDLM in which the system matrix is either similar to $\boldsymbol{J}_n(\lambda)$ or is similar to $\boldsymbol{J}$ as defined before. Let $\boldsymbol{T}$ be the observability matrix of the model and define $\boldsymbol{E}_{nr}$ as in equation (2.18). Then*

(i) any models $\mathcal{M}_1 = \{\boldsymbol{E}_{n,i}, \boldsymbol{J}_n(\lambda), \cdot, \cdot\}$ and $\mathcal{M}_1^* = \{\boldsymbol{E}_{nr}, \boldsymbol{J}, \cdot, \cdot\}$, with observability matrices $\boldsymbol{T}_1$ and $\boldsymbol{T}_1^*$ respectively, are defined as **canonical similar** *models; and*

(ii) the models $\mathcal{M}_0 = \{\boldsymbol{E}_{n,i}, \boldsymbol{J}_n(\lambda), \boldsymbol{V}_t, \boldsymbol{H}\boldsymbol{W}_t\boldsymbol{H}'\}$ and $\mathcal{M}_0^* = \{\boldsymbol{E}_{nr}, \boldsymbol{J}, \boldsymbol{V}_t, \boldsymbol{H}^*\boldsymbol{W}_t\boldsymbol{H}^{*'}\}$, where $\boldsymbol{H} = (\boldsymbol{T}_0'\boldsymbol{T}_0)^{-1}\boldsymbol{T}_0'\boldsymbol{T}$ and $\boldsymbol{H}^* = (\boldsymbol{T}_0^{*'}\boldsymbol{T}_0^*)^{-1}\boldsymbol{T}_0^{*'}\boldsymbol{T}$, respectively are defined as the **canonical equivalent** *models.*

2.7 Expert Intervention

Feed-back intervention is the action the modeller takes to assess and possibly improve the model accuracy, based on retrospecting predictive performance. Feed-forward intervention is the prospective action the modeller takes to update the model performance based upon external/expert information. Now, the focus is on forecasting and the predictive distribution is of particular interest. In practice closed models are only theoretically consistent. There will be many factors that will influence the evolution of real data. Often information will be available in qualitative form. For example a strike in the industrial market may have considerable influence on some stocks. Investors may appear rather reserved, when they anticipate a higher degree of market volatility than they expect. Information management is a crucial stage of a forecasting policy, especially when short term forecasting is the goal and when there are many kinds of information. There are three main forms of expert intervention. Here we concentrate on the more popular first two. Again the interested reader is referred to [51, chapter 11].

First assume that at time $t-1$, the available information is described by the information set $D_{t-1} = \{D_0, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{t-1}\}$, where D_0 is the initial information. Suppose that at time t all the values of the observation vector $\boldsymbol{Y}_t$ are outliers. If no intervention action is taken such that at time t it is $D_t = \{D_{t-1}, \boldsymbol{Y}_t\}$, then $\boldsymbol{Y}_t$ will dominate the forecast distribution $p(\boldsymbol{Y}_{t+1}|D_t)$, producing inadequate forecasts. A possible action would be to treat $\boldsymbol{Y}_t$ as a missing observation. Denote with I_t the additional information set based on the modeller's believes at time t. Then it is

$$I_t = \{\boldsymbol{Y}_t \text{ is missing}\}$$

and the current information set, D_t, is $D_t = \{D_{t-1}, I_t\} = D_{t-1}$. So, considering the general multivariate DLM with known variances, it is

$$\boldsymbol{m}_t = \boldsymbol{a}_t \qquad \text{and} \qquad \boldsymbol{C}_t = \boldsymbol{R}_t,$$

with $(\boldsymbol{\theta}_t|D_t) \sim \mathrm{N}[\boldsymbol{m}_t, \boldsymbol{C}_t]$. The updatings of $\boldsymbol{m}_t$ and $\boldsymbol{C}_t$ are based only on D_{t-1}, the posterior of $(\boldsymbol{\theta}_t|D_t)$ equals the prior of $(\boldsymbol{\theta}_t|D_{t-1})$, hence the forecast distribution of $(\boldsymbol{Y}_{t+1}|D_t) \equiv (\boldsymbol{Y}_{t+1}|D_{t-1})$. The above methodology can be embodied on the DLM sequential analysis if it is set $\boldsymbol{V}_t^{-1} = \mathbf{O}$. For more details on missing observations in multivariate DLMs, see Chapter 6.

Another more general form of intervention is when an additional evolution noise is considered. Consider the general multivariate DLM. Suppose that at time $t-1$ the available information is D_{t-1} and together with expert information I_t indicating a major change from $\boldsymbol{\theta}_{t-1}$ to $\boldsymbol{\theta}_t$. This can be modelled including an additional noise in the evolution equation. Suppose that the extra information set at $t-1$ is $I_t = \{\boldsymbol{h}_t, \boldsymbol{H}_t\}$, introducing the evolution noise,

$$\boldsymbol{\xi}_t \sim \mathrm{N}[\boldsymbol{h}_t, \boldsymbol{H}_t],$$

for some known moments $\boldsymbol{h}_t$, $\boldsymbol{H}_t$. Then the evolution equation becomes

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t + \boldsymbol{\xi}_t.$$

Now assuming that $\boldsymbol{\xi}_t$ is mutually independent with $\boldsymbol{\omega}_t$, the prior distribution of $(\boldsymbol{\theta}_t|D_{t-1}, I_t)$ is

$$(\boldsymbol{\theta}_t|D_{t-1}, I_t) \sim \mathrm{N}[\boldsymbol{a}_t^*, \boldsymbol{R}_t^*],$$

where $\boldsymbol{a}_t^* = \boldsymbol{a}_t + \boldsymbol{h}_t$ and $\boldsymbol{R}_t^* = \boldsymbol{R}_t + \boldsymbol{H}_t$. Also note that rewriting the evolution equation as

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t^*, \qquad \boldsymbol{\omega}_t^* \sim \mathrm{N}[\boldsymbol{h}_t, \boldsymbol{W}_t^*],$$

where $\boldsymbol{\omega}_t^* = \boldsymbol{\omega}_t + \boldsymbol{\xi}_t$ and $\boldsymbol{W}_t^* = \boldsymbol{W}_t + \boldsymbol{H}_t$ the above methodology is automatically incorporated in the DLM analysis. If there is no intervention, simply set $\boldsymbol{h}_t = \mathbf{0}$, $\boldsymbol{H}_t = \mathbf{O}$. An example of this form of intervention appears in Section 6.5.

CHAPTER 3

A Regression DLM

3.1 Introduction

In this chapter we introduce a multivariate dynamic regression model, leaving the distribution of the unknown observational variance matrix unspecified. Thus, the analysis is not restricted to the usual inverse Wishart distribution assumption. This model first appeared in [43] and is applied to the problem of computer networks security in [44, 45]. In Section 3.2 the static case is considered, which has some interest in itself. The main analysis is carried out in Section 3.3. The relationship of this model with the Common Components Model, is explored in Section 3.4 and in the next section some implementation aspects are discussed. The use of generalised inverses overcomes matrix inversion problems. A general version of the model is introduced and further developed in Section 3.6.

3.2 The Static Regression Model

Consider the model

$$\boldsymbol{Y}_t' = \boldsymbol{F}_t'\boldsymbol{\Theta} + \boldsymbol{\nu}_t', \qquad \boldsymbol{\nu}_t \sim \mathrm{N}[\mathbf{0}, \boldsymbol{\Sigma}], \tag{3.1}$$

where $\boldsymbol{Y}_t$ is an $r \times 1$ observation vector, $\boldsymbol{F}_t$ is a known $n \times 1$ design vector, $\boldsymbol{\Theta}$ an $n \times r$ parameter matrix, $\boldsymbol{\nu}_t$ an $r \times 1$ random vector and $\boldsymbol{\Sigma}$ an unknown $r \times r$ observational variance matrix.

Defining

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1' \\ \boldsymbol{Y}_2' \\ \vdots \\ \boldsymbol{Y}_t' \end{pmatrix}, \qquad \boldsymbol{F} = \begin{pmatrix} \boldsymbol{F}_1' \\ \boldsymbol{F}_2' \\ \vdots \\ \boldsymbol{F}_t' \end{pmatrix}, \qquad \boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\nu}_1' \\ \boldsymbol{\nu}_2' \\ \vdots \\ \boldsymbol{\nu}_t' \end{pmatrix},$$

where $\dim(\boldsymbol{Y}) = t \times r$, $\dim(\boldsymbol{F}) = t \times n$, and $\dim(\boldsymbol{\Psi}) = t \times r$, the model can be written compactly as

$$\boldsymbol{Y} = \boldsymbol{F}\boldsymbol{\Theta} + \boldsymbol{\Psi}, \qquad \boldsymbol{\Psi} \sim \mathrm{N}[\mathbf{O}, \boldsymbol{I}, \boldsymbol{\Sigma}].$$

$\boldsymbol{\Psi}$ follows a matrix normal distribution as discussed in [8], see Appendix B.2.

At time t, write the linear least squares estimate of $\boldsymbol{\Theta}$ as $\boldsymbol{m}_t$ and the usual estimate of $\boldsymbol{\Sigma}$ as $\boldsymbol{S}_t$. Now, define working matrices $\boldsymbol{X}_t$, $\boldsymbol{A}_t$ and $\boldsymbol{H}_t$, together with the one-step forecast error vector $\boldsymbol{e}_t$ and the residual error vector $\boldsymbol{r}_t$, by

$$\begin{aligned}
\boldsymbol{X}_t &= \sum_{i=0}^{t-1} \boldsymbol{F}_{t-i}\boldsymbol{F}'_{t-i} = \boldsymbol{F}_t\boldsymbol{F}'_t + \boldsymbol{X}_{t-1}, && \dim(\boldsymbol{X}_t) = n \times n, \\
\boldsymbol{A}_t &= \boldsymbol{X}_t^{-1}\boldsymbol{F}_t = \boldsymbol{C}_t\boldsymbol{F}_t, && \dim(\boldsymbol{A}_t) = n \times 1, \qquad (3.2)\\
\boldsymbol{H}_t &= \sum_{i=0}^{t-1} \boldsymbol{F}_{t-i}\boldsymbol{Y}'_{t-i} = \boldsymbol{F}_t\boldsymbol{Y}'_t + \boldsymbol{H}_{t-1}, && \dim(\boldsymbol{H}_t) = n \times r, \\
\boldsymbol{e}_t &= \boldsymbol{Y}_t - \boldsymbol{m}'_{t-1}\boldsymbol{F}_t, && \dim(\boldsymbol{e}_t) = r \times 1, \\
\boldsymbol{r}_t &= \boldsymbol{Y}_t - \boldsymbol{m}'_t\boldsymbol{F}_t, && \dim(\boldsymbol{r}_t) = r \times 1.
\end{aligned}$$

Lemma 3.1. *If the inverse of $\boldsymbol{X}_t$ exists for all t, then the matrix $\boldsymbol{X}_{t-1}^{-1}\boldsymbol{F}_t\boldsymbol{F}'_t\boldsymbol{X}_t^{-1}$ is a symmetric matrix.*

Proof. $\boldsymbol{X}_t$ and $\boldsymbol{X}_{t-1}$ are both symmetric, $\boldsymbol{X}_t - \boldsymbol{X}_{t-1} = \boldsymbol{F}_t\boldsymbol{F}'_t$ and

$$\boldsymbol{X}_{t-1}^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_{t-1})\boldsymbol{X}_t^{-1} = \boldsymbol{X}_{t-1}^{-1} - \boldsymbol{X}_t^{-1} = \boldsymbol{X}_t^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_{t-1})\boldsymbol{X}_{t-1}^{-1}.$$

□

Theorem 3.1. *The least squares estimate $\boldsymbol{m}_t$ of Θ may be written recurrently as $\boldsymbol{m}_t = \boldsymbol{m}_{t-1} + \boldsymbol{A}_t\boldsymbol{e}'_t$.*

Proof. Using the above lemma and the standard result that $\boldsymbol{m}_t = \boldsymbol{X}_t^{-1}\boldsymbol{H}_t$

$$\begin{aligned}
\boldsymbol{m}_t - \boldsymbol{m}_{t-1} &= \boldsymbol{X}_t^{-1}\boldsymbol{H}_t - \boldsymbol{X}_{t-1}^{-1}\boldsymbol{H}_{t-1} \\
&= \boldsymbol{X}_t^{-1}(\boldsymbol{H}_{t-1} + \boldsymbol{F}_t\boldsymbol{Y}'_t) - \boldsymbol{X}_{t-1}^{-1}\boldsymbol{H}_{t-1} \\
&= (\boldsymbol{X}_t^{-1} - \boldsymbol{X}_{t-1}^{-1})\boldsymbol{H}_{t-1} + \boldsymbol{X}_t^{-1}\boldsymbol{F}_t\boldsymbol{Y}'_t \\
&= \boldsymbol{X}_t^{-1}\boldsymbol{F}_t\boldsymbol{Y}'_t - \boldsymbol{X}_t^{-1}\boldsymbol{F}_t\boldsymbol{F}'_t\boldsymbol{X}_{t-1}^{-1}\boldsymbol{H}_{t-1} \\
&= \boldsymbol{X}_t^{-1}\boldsymbol{F}_t(\boldsymbol{Y}'_t - \boldsymbol{F}'_t\boldsymbol{X}_{t-1}^{-1}\boldsymbol{H}_{t-1}) \\
&= \boldsymbol{X}_t^{-1}\boldsymbol{F}_t(\boldsymbol{Y}'_t - \boldsymbol{F}'_t\boldsymbol{m}_{t-1}) \\
&= \boldsymbol{A}_t\boldsymbol{e}'_t.
\end{aligned}$$

□

Theorem 3.2. *At time t, the traditional variance estimate $\boldsymbol{S}_t$ of $\boldsymbol{\Sigma}$ is given by*

$$n_t\boldsymbol{S}_t = n_{t-1}\boldsymbol{S}_{t-1} + \boldsymbol{r}_t\boldsymbol{e}_t',$$

$$n_t = n_{t-1} + 1.$$

Proof. Using the traditional regression estimate $\boldsymbol{S}_t \propto \boldsymbol{Y}'(\boldsymbol{I}-\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}')\boldsymbol{Y}$ with $n_t = n_{t-1} + 1$ and as usual $n_t = t - n$, for $t > n$,

$$\begin{aligned}
n_t\boldsymbol{S}_t &= \boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{Y}'\boldsymbol{F}\boldsymbol{m}_t \\
&= \sum_{i=0}^{t-1}\boldsymbol{Y}_{t-i}\boldsymbol{Y}_{t-i}' - \boldsymbol{H}_t'\boldsymbol{m}_t \\
&= \sum_{i=0}^{t-2}\boldsymbol{Y}_{t-1-i}\boldsymbol{Y}_{t-1-i}' - \boldsymbol{H}_{t-1}'\boldsymbol{m}_{t-1} + \boldsymbol{Y}_t\boldsymbol{Y}_t' - \boldsymbol{H}_t'\boldsymbol{m}_t + \boldsymbol{H}_{t-1}'\boldsymbol{m}_{t-1} \\
&= n_{t-1}\boldsymbol{S}_{t-1} + \boldsymbol{Y}_t\boldsymbol{Y}_t' - \boldsymbol{Y}_t\boldsymbol{F}_t'\boldsymbol{m}_t - \boldsymbol{H}_{t-1}'\boldsymbol{A}_t\boldsymbol{e}_t' \\
&= n_{t-1}\boldsymbol{S}_{t-1} + (\boldsymbol{Y}_t - \boldsymbol{H}_t'\boldsymbol{A}_t)\boldsymbol{e}_t' \\
&= n_{t-1}\boldsymbol{S}_{t-1} + (\boldsymbol{Y}_t - \boldsymbol{m}_t'\boldsymbol{F}_t)\boldsymbol{e}_t' \\
&= n_{t-1}\boldsymbol{S}_{t-1} + \boldsymbol{r}_t\boldsymbol{e}_t'.
\end{aligned}$$

□

The estimate $\boldsymbol{S}_t$ of $\boldsymbol{\Sigma}$ can alternatively be written as

$$\boldsymbol{S}_t = \boldsymbol{S}_{t-1} + n_t^{-1}(\alpha_t \mathrm{E}[\boldsymbol{\nu}_t|D_t]\mathrm{E}[\boldsymbol{\nu}_t'|D_t] - \boldsymbol{S}_{t-1}),$$

where $\alpha_t = (1 - \boldsymbol{A}_t'\boldsymbol{F}_t)^{-1}$.

3.3 The Regression DLM

The regression DLM is defined by

$$\boldsymbol{Y}_t' = \boldsymbol{F}_t'\Theta_t + \boldsymbol{\nu}_t', \qquad \boldsymbol{\nu}_t \sim \mathrm{N}[\mathbf{0}, \boldsymbol{\Sigma}], \tag{3.3}$$

$$\Theta_t = \Theta_{t-1} + \boldsymbol{\Omega}_t, \qquad \boldsymbol{\Omega}_t \sim \mathrm{N}[\mathrm{O}, \boldsymbol{W}_t, \boldsymbol{\Sigma}], \tag{3.3'}$$

where Θ_t is a $n \times r$ state or parameter matrix, $\boldsymbol{\Omega}_t$ an $n \times r$ random matrix, $\boldsymbol{W}_t$ a known $n \times n$ evolution matrix and all the remaining components are as defined in Section 3.2. Further, the noise terms $\boldsymbol{\nu}_t$ and $\boldsymbol{\Omega}_t$ are mutually and internally independent. The evolution equation (3.3′) models the state matrix as a random walk and as locally constant. $\boldsymbol{W}_t$ is specified via the discounting approach explained in Chapter 2

$$\boldsymbol{W}_t = \frac{1-\delta}{\delta}\boldsymbol{C}_{t-1}, \tag{3.4}$$

where δ is a discount factor and $\boldsymbol{S}_{t-1} \otimes \boldsymbol{C}_{t-1} = \mathrm{V}[\mathrm{vec}(\Theta_{t-1})|D_{t-1}, \boldsymbol{\Sigma} = \boldsymbol{S}_{t-1}]$, for $\boldsymbol{S}_{t-1}$, $\boldsymbol{C}_{t-1}$ having been calculated at time $t-1$.

This model is referred to as NDRDLM (Normal Discount Regression DLM).

In order to estimate the parameters of model (3.3), (3.3′), we employ DWR (Discount Weighted Regression).

Multivariate DWR estimates $\Theta_t = \Theta$ as $\boldsymbol{m}_t$, based upon the minimisation of the discounted sum of squares

$$S_\delta(\Theta) = \sum_{i=0}^{t-1} \delta^i (\boldsymbol{Y}_{t-i}' - \boldsymbol{F}_{t-i}'\Theta)(\boldsymbol{Y}_{t-i}' - \boldsymbol{F}_{t-i}'\Theta)',$$

where δ is a discount factor ($0 < \delta \leq 1$).

Following Ameen and Harrison ([2]) we adopt a forecast function $\boldsymbol{f}_t(k) = \boldsymbol{m}_t'\boldsymbol{F}_{t+k}$. NDRDLM and DWR are equivalent in the sense that both provide the same forecast function (see [51, page 207]). Moreover, assuming a vague prior $C_0^{-1} \approx \mathbf{O}$ for the NDRDLM, DWR converges to it ([51, chapter 6]). Thus, we can use the DWR to estimate the parameters of the NDRDLM. Now

$$\sqrt{\delta^i}\boldsymbol{Y}'_{t-i} = \sqrt{\delta^i}\boldsymbol{F}'_{t-i}\Theta + \sqrt{\delta^i}\boldsymbol{\nu}'_{t-i}, \qquad \boldsymbol{\nu}_{t-i} \sim \mathrm{N}[\mathbf{0}, \Sigma] \tag{3.5}$$

and since DWR is a weighted regression the least squares estimates are identical to those from static model (model (3.5)) that may be written as

$$\boldsymbol{Y}^* = \boldsymbol{F}^*\Theta + \Psi^*, \qquad \Psi^* \sim \mathrm{N}[\mathbf{O}, \mathrm{diag}\{\delta^{t-1}, \delta^{t-2}, \dots, 1\}, \Sigma],$$

where

$$\boldsymbol{Y}^* = \begin{pmatrix} \sqrt{\delta^{t-1}}\boldsymbol{Y}'_1 \\ \sqrt{\delta^{t-2}}\boldsymbol{Y}'_2 \\ \vdots \\ \boldsymbol{Y}'_t \end{pmatrix}, \quad \boldsymbol{F}^* = \begin{pmatrix} \sqrt{\delta^{t-1}}\boldsymbol{F}'_1 \\ \sqrt{\delta^{t-2}}\boldsymbol{F}'_2 \\ \vdots \\ \boldsymbol{F}'_t \end{pmatrix}, \quad \Psi^* = \begin{pmatrix} \sqrt{\delta^{t-1}}\boldsymbol{\nu}'_1 \\ \sqrt{\delta^{t-2}}\boldsymbol{\nu}'_2 \\ \vdots \\ \boldsymbol{\nu}'_t \end{pmatrix}.$$

Retaining the previous definitions of $\boldsymbol{A}_t$, $\boldsymbol{e}_t$, and $\boldsymbol{r}_t$, redefine the working matrices

$$\boldsymbol{X}_t = \sum_{i=0}^{t-1} \delta^i \boldsymbol{F}_{t-i}\boldsymbol{F}'_{t-i} = \boldsymbol{F}_t\boldsymbol{F}'_t + \delta\boldsymbol{X}_{t-1}, \qquad \dim(\boldsymbol{X}_t) = n \times n,$$

$$\boldsymbol{H}_t = \sum_{i=0}^{t-1} \delta^i \boldsymbol{F}_{t-i}\boldsymbol{Y}'_{t-i} = \boldsymbol{F}_t\boldsymbol{Y}'_t + \delta\boldsymbol{H}_{t-1}, \qquad \dim(\boldsymbol{H}_t) = n \times r.$$

Lemma 3.2. *If the inverse of $\boldsymbol{X}_t$ exists for all t, then the matrix $\boldsymbol{X}_{t-1}^{-1}\boldsymbol{F}_t\boldsymbol{F}'_t\boldsymbol{X}_t^{-1}$ is a symmetric matrix.*

Proof. $\boldsymbol{X}_t$ and $\boldsymbol{X}_{t-1}$ are both symmetric, $\boldsymbol{X}_t - \delta\boldsymbol{X}_{t-1} = \boldsymbol{F}_t\boldsymbol{F}'_t$ and

$$\boldsymbol{X}_{t-1}^{-1}(\boldsymbol{X}_t - \delta\boldsymbol{X}_{t-1})\boldsymbol{X}_t^{-1} = \boldsymbol{X}_{t-1}^{-1} - \delta\boldsymbol{X}_t^{-1} = \boldsymbol{X}_t^{-1}(\boldsymbol{X}_t - \delta\boldsymbol{X}_{t-1})\boldsymbol{X}_{t-1}^{-1}.$$

□

Theorem 3.3. *The least squares estimate of* Θ_t *is* $\boldsymbol{m}_t = \boldsymbol{m}_{t-1} + \boldsymbol{A}_t\boldsymbol{e}_t'$.

Proof. Using the above lemma and the standard result that $\boldsymbol{m}_t = \boldsymbol{X}_t^{-1}\boldsymbol{H}_t$

$$\begin{aligned}
\boldsymbol{m}_t - \boldsymbol{m}_{t-1} &= \boldsymbol{X}_t^{-1}\boldsymbol{H}_t - \boldsymbol{X}_{t-1}^{-1}\boldsymbol{H}_{t-1} \\
&= \boldsymbol{X}_t^{-1}(\delta\boldsymbol{H}_{t-1} + \boldsymbol{F}_t\boldsymbol{Y}_t') - \boldsymbol{X}_{t-1}^{-1}\boldsymbol{H}_{t-1} \\
&= (\delta\boldsymbol{X}_t^{-1} - \boldsymbol{X}_{t-1}^{-1})\boldsymbol{H}_{t-1} + \boldsymbol{X}_t^{-1}\boldsymbol{F}_t\boldsymbol{Y}_t' \\
&= \boldsymbol{X}_t^{-1}\boldsymbol{F}_t\boldsymbol{Y}_t' - \boldsymbol{X}_t^{-1}\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{X}_{t-1}^{-1}\boldsymbol{H}_{t-1} \\
&= \boldsymbol{X}_t^{-1}\boldsymbol{F}_t(\boldsymbol{Y}_t' - \boldsymbol{F}_t'\boldsymbol{X}_{t-1}^{-1}\boldsymbol{H}_{t-1}) \\
&= \boldsymbol{X}_t^{-1}\boldsymbol{F}_t(\boldsymbol{Y}_t' - \boldsymbol{F}_t'\boldsymbol{m}_{t-1}) \\
&= \boldsymbol{A}_t\boldsymbol{e}_t'.
\end{aligned}$$

$\square$

Since $\boldsymbol{\Sigma}$ is constant over t, its estimate $\boldsymbol{S}_t$ will be calculable from Theorem 3.2.

Let D_0 be the initial information and given D_0 and $\boldsymbol{\Sigma} = \boldsymbol{S}_0$ for a point initial estimate of $\boldsymbol{\Sigma}$,

$$(\Theta_0|D_0, \boldsymbol{\Sigma} = \boldsymbol{S}_0) \sim \mathrm{N}[\boldsymbol{m}_0, \boldsymbol{C}_0, \boldsymbol{S}_0].$$

The following theorem summarises all the above analysis.

Theorem 3.4. *For the regression DLM (3.3), (3.3') with the evolution variance as in (3.4), assuming that the inverse of* $\boldsymbol{X}_t$ *exists, the following results hold*

(a) Posterior at $t-1$*:*

$$(\Theta_{t-1}|D_{t-1}, \boldsymbol{\Sigma} = \boldsymbol{S}_{t-1}) \sim N[\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}, \boldsymbol{S}_{t-1}],$$

for some known quantities $\boldsymbol{m}_{t-1}$, $\boldsymbol{C}_{t-1}$, $\boldsymbol{S}_{t-1}$.

(b) Prior at t:

$$(\boldsymbol{\Theta}_t | D_{t-1}, \boldsymbol{\Sigma} = \boldsymbol{S}_{t-1}) \sim N[\boldsymbol{a}_t, \boldsymbol{R}_t, \boldsymbol{S}_{t-1}],$$

where

$$\boldsymbol{a}_t = \boldsymbol{m}_{t-1} \qquad \textit{and} \qquad \boldsymbol{R}_t = \boldsymbol{C}_{t-1}/\delta.$$

(c) One-step forecast:

$$(\boldsymbol{Y}_t | D_{t-1}, \boldsymbol{\Sigma} = \boldsymbol{S}_{t-1}) \sim N[\boldsymbol{f}_t, \boldsymbol{Q}_t],$$

where

$$\boldsymbol{f}_t' = \boldsymbol{F}_t'\boldsymbol{a}_t \qquad \textit{and} \qquad \boldsymbol{Q}_t = (\boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + 1)\boldsymbol{S}_{t-1}.$$

(d) Posterior at t:

$$(\boldsymbol{\Theta}_t | D_t, \boldsymbol{\Sigma} = \boldsymbol{S}_t) \sim N[\boldsymbol{m}_t, \boldsymbol{C}_t, \boldsymbol{S}_t],$$

where

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t\boldsymbol{e}_t', \quad \boldsymbol{C}_t = \frac{1}{\delta}\left[\boldsymbol{I} - \frac{\boldsymbol{C}_{t-1}\boldsymbol{F}_t\boldsymbol{F}_t'}{\delta + \boldsymbol{F}_t'\boldsymbol{C}_{t-1}\boldsymbol{F}_t}\right]\boldsymbol{C}_{t-1},$$

$$n_t\boldsymbol{S}_t = n_{t-1}\boldsymbol{S}_{t-1} + \boldsymbol{r}_t\boldsymbol{e}_t',$$

with

$$\boldsymbol{A}_t = \boldsymbol{C}_t\boldsymbol{F}_t, \quad \boldsymbol{e}_t = \boldsymbol{Y}_t - \boldsymbol{f}_t, \quad \boldsymbol{r}_t = \boldsymbol{Y}_t - \boldsymbol{m}_t'\boldsymbol{F}_t, \quad n_t = n_{t-1} + 1.$$

Proof. The proof is by induction. Assume that (a) is true. By applying the transformation

$$\boldsymbol{Y}_t^* = \mathrm{vec}(\boldsymbol{Y}_t'), \qquad \boldsymbol{F}_t^* = \boldsymbol{I}_r \otimes \boldsymbol{F}_t, \qquad \boldsymbol{\theta}_t^* = \mathrm{vec}(\boldsymbol{\Theta}_t),$$

$$\boldsymbol{\nu}_t^* = \mathrm{vec}(\boldsymbol{\nu}_t'), \qquad \boldsymbol{G}_t^* = \boldsymbol{I}_r \otimes \boldsymbol{G}_t, \qquad \boldsymbol{\omega}_t^* = \mathrm{vec}(\boldsymbol{\Omega}_t),$$

model (3.3), (3.3′) can be written as

$$Y_t^* = F_t^{*\prime}\theta_t^* + \nu_t^*, \qquad \nu_t^* \sim \mathrm{N}[\mathbf{0}, \Sigma],$$
$$\theta_t^* = G_t^*\theta_{t-1}^* + \omega_t^*, \qquad \omega_t^* \sim \mathrm{N}[\mathbf{0}, \Sigma \otimes W_t],$$

where $W_t = (\delta^{-1} - 1)C_{t-1}$.

Then conditional on $\Sigma = S_{t-1}$ the distributions of (b), (c) of the theorem are simply the results of the known variance case, given in Theorem 2.3 (page 17).

The recurrences of m_t and S_t of (d) are essentially derived from Theorems 3.3 and 3.2 respectively. A_t comes from the definition of page 44. To prove the updating of C_t we use Lemma 3.2. So we have

$$\begin{aligned}
C_{t-1} &= C_t(\delta I + F_tF_t'C_{t-1}) \Longrightarrow \\
C_tF_t &= (\delta + F_t'C_{t-1}F_t)^{-1}C_{t-1}F_t \Longrightarrow \\
C_{t-1} - \delta C_t &= (\delta + F_t'C_{t-1}F_t)^{-1}C_{t-1}F_tF_t'C_{t-1} \Longrightarrow \\
C_t &= \frac{1}{\delta}\left[I - \frac{C_{t-1}F_tF_t'}{\delta + F_t'C_{t-1}F_t}\right]C_{t-1},
\end{aligned}$$

which establishes (d). The proof is completed by noticing that (a) is initially true for $t = 1$. □

Note that no matrix inversion is required, hence the method is fast and computationally efficient.

More details about DWR and its relation with DLMs are to be found in [1, chapters 2,3] (for the univariate case) and in [48, chapter 6] (for the multivariate case with known variances).

3.4 Relationship with the Common Components Model

If we assume an inverse Wishart distribution for the unknown variance matrix Σ, model (3.3), (3.3′) would be a special case of a discounted CCM (see Section 2.3.2). In particular in the CCM, setting $\boldsymbol{G}_t = \boldsymbol{I}$, $\boldsymbol{W}_t = (\delta^{-1}-1)\boldsymbol{C}_{t-1}$ and $V_t = 1$, we nearly get the NDRDLM. The only difference being the modelling of Σ. So these models are very similar and the relationship of the CCM with the DWR can be used to explore the relationship of the NDRDLM and the CCM. First we show that the estimate, $\boldsymbol{S}_t$ of Theorem 3.2, used in Section 3.3, coincides with the usual estimate of the CCM.

To see this write

$$\begin{aligned} n_t\boldsymbol{S}_t &= n_{t-1}\boldsymbol{S}_{t-1} + \boldsymbol{r}_t\boldsymbol{e}_t' \\ &= n_{t-1}\boldsymbol{S}_{t-1} + \boldsymbol{e}_t(1 - \boldsymbol{A}_t'\boldsymbol{F}_t)\boldsymbol{e}_t' \\ &= n_{t-1}\boldsymbol{S}_{t-1} + \boldsymbol{e}_t\boldsymbol{e}_t'/Q_t. \end{aligned}$$

The next theorem gives the relationship for $\boldsymbol{m}_t$, $\boldsymbol{C}_t$ of the two models.

Theorem 3.5. *For the CCM and the DWR the following results hold*

(a) Assuming a vague prior $\boldsymbol{C}_0^{-1} \approx \mathbf{O}$, the estimates, $\boldsymbol{m}_t$, $\boldsymbol{C}_t$, produced by each model coincide.

(b) Assuming that $\lim_{t\to\infty} \boldsymbol{C}_t = \boldsymbol{C}$ exists and is a non-singular matrix, the limiting values of $\boldsymbol{m}_t$, $\boldsymbol{C}_t$, produced by each model coincide.

Proof. First we prove (a). Consider the CCM and write down the usual equations

$$\begin{aligned} \boldsymbol{C}_t^{-1} &= \boldsymbol{R}_t^{-1} + \boldsymbol{F}_t\boldsymbol{F}_t'|V_t \\ &= \delta\boldsymbol{C}_{t-1}^{-1} + \boldsymbol{F}_t\boldsymbol{F}_t' \qquad (3.6) \\ \boldsymbol{C}_t^{-1}\boldsymbol{m}_t &= \delta\boldsymbol{C}_{t-1}^{-1}\boldsymbol{m}_{t-1} + \boldsymbol{F}_t\boldsymbol{Y}_t'. \qquad (3.7) \end{aligned}$$

Applying equation (3.6) recursively we have

$$\boldsymbol{C}_t^{-1} = \sum_{i=0}^{t-1} \delta^i \boldsymbol{F}_{t-i}\boldsymbol{F}_{t-i}' + \delta^t \boldsymbol{C}_0^{-1}. \qquad (3.8)$$

Similarly from (3.7), using (3.8) we obtain

$$\boldsymbol{m}_t = \boldsymbol{C}_t \left(\sum_{i=0}^{t-1} \delta^i \boldsymbol{F}_{t-i}\boldsymbol{Y}_{t-i}' + \delta^t \boldsymbol{C}_0^{-1}\boldsymbol{m}_0 \right). \qquad (3.9)$$

Assumption $\boldsymbol{C}_0^{-1} \approx \mathbf{O}$ implies that $\boldsymbol{C}_t$ and $\boldsymbol{m}_t$ of equations (3.8) and (3.9) are the same with the respective of DWR.

The proof of (b) is trivial by noticing that $\lim_{t\to\infty} \delta^t \boldsymbol{C}_0^{-1} = \mathbf{O}$ and $\lim_{t\to\infty} \delta^t \boldsymbol{C}_0^{-1}\boldsymbol{m}_0 = \mathbf{0}$ as $t \to \infty$. □

If we assume that $\lim_{t\to\infty} \boldsymbol{F}_t = \boldsymbol{F}$ exists, then we have the following limiting results (necessarily the same for both CCM and DWR).

$$\boldsymbol{R} = \boldsymbol{C}/\delta,$$

$$\boldsymbol{A} = \boldsymbol{C}\boldsymbol{F},$$

$$\boldsymbol{Q} = (\boldsymbol{F}'\boldsymbol{R}\boldsymbol{F} + 1)\boldsymbol{\Sigma},$$

where $\boldsymbol{R} = \lim_{t\to\infty} \boldsymbol{R}_t$, $\boldsymbol{C} = \lim_{t\to\infty} \boldsymbol{C}_t$, $\boldsymbol{A} = \lim_{t\to\infty} \boldsymbol{A}_t$, $\boldsymbol{Q} = \lim_{t\to\infty} \boldsymbol{Q}_t$, $\boldsymbol{\Sigma} = \lim_{t\to\infty} \boldsymbol{S}_t$. This is immediate from Theorem 3.5 and $\lim_{t\to\infty} \boldsymbol{S}_t = \boldsymbol{\Sigma}$.

The consequence of the above analysis is that when a vague prior $\boldsymbol{C}_0^{-1}$ is used the DWR can be used to calculate the estimates of the NDRDLM. Furthermore, both these models have the same limiting behaviour not depending on $\boldsymbol{C}_0^{-1}$.

3.5 Inversion Problems

In Section 3.3 we proved the updatings for m_t and S_t, in Theorems 3.3 and 3.2. However, both theorems assume the existence of the matrix X_t^{-1} for all t. This may not always exist, for example when F_t does not depend on t. Let $F_t = X$, an $n \times 1$ known constant vector for all t. Then

$$X_t = \frac{1-\delta^t}{1-\delta} X X',$$

with $|X_t| = 0$, for all t, since X is a vector. Clearly, in such a case we cannot proceed as in Section 3.3.

Estimates m_t and S_t can be derived using generalised inverses. It is not, however, then possible in general to retain the neat formulae of Theorems 3.3 and 3.2. The quantities m_t, S_t will always exist, but their calculation will necessarily involve F^*, Y^*. The following theorem introduces a weaker assumption than the non-singularity of X_t that will allow the same recurrences.

Let X_t^+ be the Moore-Penrose inverse of X_t for all t, (see Appendix A.3). Redefining $C_t = X_t^+$ and $A_t = X_t^+ F_t$ we have

Theorem 3.6. *In the framework of Section 3.3, if* *rank*(X_{t-1}^+, X_t^+) =*rank*(X_{t-1}) =*rank*(X_t)*, then*

$$m_t = m_{t-1} + A_t e_t'.$$

Proof. The hypothesis implies that both linear systems $X_{t-1}^+ X = X_t^+$ and $X_t^+ X = X_{t-1}^+$ are consistent. Thus from Theorem A.3

$$X_{t-1}^+ X_{t-1} X_t^+ = X_t^+ \qquad \text{and} \qquad X_t^+ X_t X_{t-1}^+ = X_{t-1}^+.$$

Now

$$\begin{aligned} \boldsymbol{X}_{t-1}^{+} - \delta \boldsymbol{X}_t^{+} &= \boldsymbol{X}_t^{+}(\boldsymbol{X}_t - \delta \boldsymbol{X}_{t-1})\boldsymbol{X}_{t-1}^{+} \\ &= \boldsymbol{X}_t^{+}\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{X}_{t-1}^{+}. \end{aligned}$$

So using $\boldsymbol{m}_t = \boldsymbol{X}_t^{+}\boldsymbol{H}_t$

$$\begin{aligned} \boldsymbol{m}_t - \boldsymbol{m}_{t-1} &= \boldsymbol{X}_t^{+}\boldsymbol{H}_t - \boldsymbol{X}_{t-1}^{+}\boldsymbol{H}_{t-1} \\ &= \boldsymbol{X}_t^{+}(\delta \boldsymbol{H}_{t-1} + \boldsymbol{F}_t\boldsymbol{Y}_t') - \boldsymbol{X}_{t-1}^{+}\boldsymbol{H}_{t-1} \\ &= (\delta \boldsymbol{X}_t^{+} - \boldsymbol{X}_{t-1}^{+})\boldsymbol{H}_{t-1} + \boldsymbol{X}_t^{+}\boldsymbol{F}_t\boldsymbol{Y}_t' \\ &= \boldsymbol{X}_t^{+}\boldsymbol{F}_t\boldsymbol{Y}_t' - \boldsymbol{X}_t^{+}\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{X}_{t-1}^{+}\boldsymbol{H}_{t-1} \\ &= \boldsymbol{X}_t^{+}\boldsymbol{F}_t(\boldsymbol{Y}_t' - \boldsymbol{F}_t'\boldsymbol{X}_{t-1}^{+}\boldsymbol{H}_{t-1}) \\ &= \boldsymbol{X}_t^{+}\boldsymbol{F}_t(\boldsymbol{Y}_t' - \boldsymbol{F}_t'\boldsymbol{m}_{t-1}) \\ &= \boldsymbol{A}_t\boldsymbol{e}_t'. \end{aligned}$$

□

This is a key result. The recurrences of $\boldsymbol{S}_t$, $\boldsymbol{C}_t$ are as in Theorems 3.2, 3.4. The proof of the latter is trivial, employing Theorem 3.6 and using $\boldsymbol{C}_t = \boldsymbol{X}_t^{+}$. Note that from Corollary A.1 it follows that if $\boldsymbol{X}^{-1}$ exists, $\boldsymbol{X}^{+} = \boldsymbol{X}^{-1}$ and the updatings of Sections 3.3 and 3.5 coincide.

In the beginning of this section it was stated that if $\boldsymbol{F}_t$ is constant $\boldsymbol{X}_t$ will be singular and the use of Moore-Penrose inverses were proposed. Note, however, that the case of a constant design vector is of little practical importance. Model (3.3), (3.3′) with $\boldsymbol{F}_t = \boldsymbol{X}$ ($\forall t > 0$) is always unobservable and by setting $\boldsymbol{\psi}_t' = \boldsymbol{X}'\Theta_t$ it can be reduced to the first-order r-dimensional polynomial DLM, namely

$$\boldsymbol{Y}_t = \boldsymbol{\psi}_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim \mathrm{N}[\mathbf{0}, \Sigma],$$

$$\boldsymbol{\psi}_t = \boldsymbol{\psi}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathrm{N}[\mathbf{0}, \boldsymbol{X}'\boldsymbol{W}_t\boldsymbol{X}\Sigma].$$

Of course $\boldsymbol{X}_t$ may well be singular, although $\boldsymbol{F}_t$ changes with time. If this is the case, the analysis developed in this section may be applied.

3.6 The General Regression DLM

In this section we develop a more general methodology when the series of interest, $\boldsymbol{Y}_t$, is a matrix. The matrix version of model (3.3), (3.3′) is defined by

$$\boldsymbol{Y}'_t = \boldsymbol{F}'_t\Theta_t + \boldsymbol{\nu}'_t, \qquad \boldsymbol{\nu}'_t \sim \mathrm{N}[\mathbf{O}, \boldsymbol{V}_t, \Sigma], \tag{3.10}$$

$$\Theta_t = \Theta_{t-1} + \boldsymbol{\Omega}_t, \qquad \boldsymbol{\Omega}_t \sim \mathrm{N}[\mathbf{O}, \boldsymbol{W}_t, \Sigma], \tag{3.10′}$$

where $\boldsymbol{Y}_t$ is an $r \times m$ observation matrix, $\boldsymbol{F}_t$ a known $n \times m$ design matrix, Θ_t an $n \times r$ state matrix, $\boldsymbol{\nu}_t$ an $r \times m$ random matrix, $\boldsymbol{V}_t$ a known $m \times m$ variance matrix, Σ an unknown $r \times r$ variance matrix, $\boldsymbol{\Omega}_t$ an $n \times r$ random matrix, and $\boldsymbol{W}_t$ an $n \times n$ variance matrix, specified by (3.4). Often practitioners will use $\boldsymbol{V}_t = \boldsymbol{I}$. Retain the definitions of $\boldsymbol{e}_t$ and $\boldsymbol{r}_t$ of Section 3.2 and redefine

$$\begin{aligned}
\boldsymbol{X}_t &= \sum_{i=0}^{t-1} \delta^i \boldsymbol{F}_{t-i}\boldsymbol{V}_{t-i}^{-1}\boldsymbol{F}'_{t-i}, && \dim(\boldsymbol{X}_t) = n \times n, \\
\boldsymbol{A}_t &= \boldsymbol{C}_t\boldsymbol{F}_t\boldsymbol{V}_t^{-1}, && \dim(\boldsymbol{A}_t) = n \times m, \\
\boldsymbol{H}_t &= \sum_{i=0}^{t-1} \delta^i \boldsymbol{F}_{t-i}\boldsymbol{V}_{t-i}^{-1}\boldsymbol{Y}'_{t-i}, && \dim(\boldsymbol{H}_t) = n \times r.
\end{aligned} \tag{3.11}$$

The DWR methodology is employed for estimating the parameters of model (3.10), (3.10′) as explained in Section 3.3.

Now, Lemma 3.2 still remains with a small modification.

Lemma 3.3. *If the inverse of X_t exists for all t, then the matrix $X_{t-1}^{-1}F_tV_t^{-1}F_t'X_t^{-1}$ is a symmetric matrix.*

Proof. X_t and X_{t-1} are both symmetric, $X_t - \delta X_{t-1} = F_tV_t^{-1}F_t'$ and

$$X_{t-1}^{-1}(X_t - \delta X_{t-1})X_t^{-1} = X_{t-1}^{-1} - \delta X_t^{-1} = X_t^{-1}(X_t - \delta X_{t-1})X_{t-1}^{-1}.$$

□

Theorem 3.7. *The least squares estimate of Θ_t is $m_t = m_{t-1} + A_te_t'$.*

Proof. Using the above lemma and the standard result that $m_t = X_t^{-1}H_t$

$$\begin{aligned}
m_t - m_{t-1} &= X_t^{-1}H_t - X_{t-1}^{-1}H_{t-1} \\
&= X_t^{-1}(\delta H_{t-1} + F_tV_t^{-1}Y_t') - X_{t-1}^{-1}H_{t-1} \\
&= (\delta X_t^{-1} - X_{t-1}^{-1})H_{t-1} + X_t^{-1}F_tV_t^{-1}Y_t' \\
&= X_t^{-1}F_tV_t^{-1}Y_t' - X_t^{-1}F_tV_t^{-1}F_t'X_{t-1}^{-1}H_{t-1} \\
&= X_t^{-1}F_tV_t^{-1}(Y_t' - F_t'X_{t-1}^{-1}H_{t-1}) \\
&= X_t^{-1}F_tV_t^{-1}(Y_t' - F_t'm_{t-1}) \\
&= A_te_t'.
\end{aligned}$$

□

Theorem 3.8. *The variance estimate S_t of Σ at time t, is given by*

$$n_tS_t = n_{t-1}S_{t-1} + r_tV_t^{-1}e_t'$$

$$n_t = n_{t-1} + 1.$$

Proof. Since Σ is constant and not depending on δ, it can be derived by the static regression. Thus setting $\boldsymbol{W}_t = \mathbf{O}$ ($\delta = 1$) and using as before the traditional regression estimate $\boldsymbol{S}_t$, with $n_t = n_{t-1} + 1$ and $n_t = t - n$, for $t > n$,

$$\begin{aligned} n_t \boldsymbol{S}_t &= \sum_{i=0}^{t-1} \boldsymbol{Y}_{t-i} \boldsymbol{V}_{t-i}^{-1} \boldsymbol{Y}'_{t-i} - \boldsymbol{H}'_t \boldsymbol{m}_t \\ &= n_{t-1} \boldsymbol{S}_{t-1} + (\boldsymbol{Y}_t \boldsymbol{V}_t^{-1} - \boldsymbol{H}'_t \boldsymbol{A}_t) e'_t \\ &= n_{t-1} \boldsymbol{S}_{t-1} + (\boldsymbol{Y}_t \boldsymbol{V}_t^{-1} - \boldsymbol{m}'_t \boldsymbol{F}_t \boldsymbol{V}_t^{-1}) e'_t \\ &= n_{t-1} \boldsymbol{S}_{t-1} + \boldsymbol{r}_t \boldsymbol{V}_t^{-1} \boldsymbol{e}'_t. \end{aligned}$$

□

Let D_0 be the initial information and given D_0 and $\Sigma = \boldsymbol{S}_0$ for a point initial estimate of Σ,

$$(\Theta_0 | D_0, \Sigma = \boldsymbol{S}_0) \sim \mathrm{N}[\boldsymbol{m}_0, \boldsymbol{C}_0, \boldsymbol{S}_0].$$

for some known quantities $\boldsymbol{m}_0$ and $\boldsymbol{C}_0$.

Theorem 3.9. *For the regression DLM (3.10), (3.10′) with the evolution variance as in (3.4), assuming that the inverse of $\boldsymbol{X}_t$ exists, the following results hold*

(a) Posterior at $t-1$:

$$(\Theta_{t-1} | D_{t-1}, \Sigma = \boldsymbol{S}_{t-1}) \sim N[\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}, \boldsymbol{S}_{t-1}],$$

for some known quantities $\boldsymbol{m}_{t-1}$, $\boldsymbol{C}_{t-1}$, $\boldsymbol{S}_{t-1}$.

(b) Prior at t:

$$(\Theta_t | D_{t-1}, \Sigma = \boldsymbol{S}_{t-1}) \sim N[\boldsymbol{a}_t, \boldsymbol{R}_t, \boldsymbol{S}_{t-1}],$$

where

$$a_t = m_{t-1} \qquad \textit{and} \qquad R_t = C_{t-1}/\delta.$$

(c) One-step forecast:

$$(Y'_t | D_{t-1}, \Sigma = S_{t-1}) \sim N[f'_t, Q_t, S_{t-1}],$$

where

$$f'_t = F'_t a_t \qquad \textit{and} \qquad Q_t = F'_t R_t F_t + V_t.$$

(d) Posterior at t:

$$(\Theta_t | D_t, \Sigma = S_t) \sim N[m_t, C_t, S_t],$$

where

$$m_t = a_t + A_t e'_t,$$

$$C_t = \frac{1}{\delta}\Big[I - C_{t-1}F_t(\delta I + V_t^{-1}F'_t C_{t-1}F_t)^{-1}V_t^{-1}F'_t\Big]C_{t-1},$$

$$n_t S_t = n_{t-1}S_{t-1} + r_t V_t^{-1} e'_t,$$

with

$$A_t = C_t F_t V_t^{-1}, \quad e_t = Y_t - f_t, \quad r_t = Y_t - m'_t F_t, \quad n_t = n_{t-1} + 1.$$

Proof. The proof is by induction. Assume that (a) is true. By applying the transformation

$$Y_t^* = \mathrm{vec}(Y'_t), \qquad F_t^* = I_r \otimes F_t, \qquad \theta_t^* = \mathrm{vec}(\Theta_t),$$

$$\nu_t^* = \mathrm{vec}(\nu'_t), \qquad G_t^* = I_r \otimes G_t, \qquad \omega_t^* = \mathrm{vec}(\Omega_t),$$

model (3.10), (3.10′) can be written as

$$Y_t^* = F_t^{*'}\theta_t^* + \nu_t^*, \qquad \nu_t^* \sim \mathrm{N}[0, \Sigma \otimes V_t],$$

$$\theta_t^* = G_t^*\theta_{t-1}^* + \omega_t^*, \qquad \omega_t^* \sim \mathrm{N}[0, \Sigma \otimes W_t],$$

where $\boldsymbol{W}_t = (\delta^{-1} - 1)\boldsymbol{C}_{t-1}$.

Then conditional on $\Sigma = \boldsymbol{S}_{t-1}$ the distributions of (b), (c) of the theorem are simply the results of the known variance case, given in Theorem 2.3 (page 17).

The recurrences of $\boldsymbol{m}_t$ and $\boldsymbol{S}_t$ of (d) are essentially derived from Theorems 3.7 and 3.8 respectively. $\boldsymbol{A}_t$ comes from the definition of page 55. To prove the updating of $\boldsymbol{C}_t$ we use Lemma 3.3. So we have

$$\begin{aligned}
\boldsymbol{C}_{t-1} &= \boldsymbol{C}_t(\delta \boldsymbol{I} + \boldsymbol{F}_t \boldsymbol{V}_t^{-1} \boldsymbol{F}_t' \boldsymbol{C}_{t-1}) \Longrightarrow \\
\boldsymbol{C}_t \boldsymbol{F}_t &= \boldsymbol{C}_{t-1} \boldsymbol{F}_t (\delta \boldsymbol{I} + \boldsymbol{V}_t^{-1} \boldsymbol{F}_t' \boldsymbol{C}_{t-1} \boldsymbol{F}_t)^{-1} \Longrightarrow \\
\boldsymbol{C}_{t-1} - \delta \boldsymbol{C}_t &= \boldsymbol{C}_{t-1} \boldsymbol{F}_t (\delta \boldsymbol{I} + \boldsymbol{V}_t^{-1} \boldsymbol{F}_t' \boldsymbol{C}_{t-1} \boldsymbol{F}_t)^{-1} \boldsymbol{V}_t^{-1} \boldsymbol{F}_t' \boldsymbol{C}_{t-1} \Longrightarrow \\
\boldsymbol{C}_t &= \frac{1}{\delta}\Big[\boldsymbol{I} - \boldsymbol{C}_{t-1} \boldsymbol{F}_t (\delta \boldsymbol{I} + \boldsymbol{V}_t^{-1} \boldsymbol{F}_t' \boldsymbol{C}_{t-1} \boldsymbol{F}_t)^{-1} \boldsymbol{V}_t^{-1} \boldsymbol{F}_t'\Big] \boldsymbol{C}_{t-1},
\end{aligned}$$

which establishes (d). The proof is completed by noticing that (a) is initially true for $t = 1$. □

The above theorem is subject to the non-singularity of $\boldsymbol{X}_t$. If $\boldsymbol{X}_t$ is singular for some t, then under the assumptions of Theorem 3.6, action similar to Section 3.5 can be taken. In such a case defining $\boldsymbol{C}_t = \boldsymbol{X}_t^+$ all the results of Theorems 3.7, 3.8, 3.9 remain the same, the proofs being in line with Section 3.5.

The variance matrix $\boldsymbol{V}_t$ is assumed known. If it is unknown the regression DLM analysis does not appear to be tractable as stated in [34].

Similar results apply as in Section 3.4 concerning the equivalence of the General Regression DLM and the matrix form of the CCM, as introduced in [34].

CHAPTER 4

New Distributional Directions to the DLM

4.1 Introduction

In this chapter we extend the existing inverse Wishart and T distributions to incorporate a matrix of degrees of freedom so that missing observation analysis and intervention are possible. In Sections 4.2 and 4.3 the new distributions are introduced and some of their properties are discussed. Section 4.4 develops the main results for the CCM (common components model) using the new distributions. These results follow [40]. Section 4.4 deals with retrospective analysis, including the deletion of observations. The following section develops the relevant reference analysis, allowing for any missing data and Section 4.7 updates an approximation method for the general multivariate model, proposed by Barbosa and Harrison, [4], using again the new distributions.

4.2 Generalized Wishart and Inverse Wishart Distributions

First a reminder of the inverse Wishart distribution is given. A detailed discussion, including historical references, can be found in ([32, chapter 5]) or ([14, chapter 3]). Let $\boldsymbol{\Sigma}$ be an $r \times r$ SPD (symmetric positive definite) random matrix $\boldsymbol{\Sigma}$, $\boldsymbol{R}$ an $r \times r$ SPD matrix, and k a positive scalar. Then the inverse Wishart distribution with k degrees of freedom is defined by

$$p(\boldsymbol{\Sigma}) = c_0|\boldsymbol{R}|^{(k-r-1)/2}|\boldsymbol{\Sigma}|^{-k/2}\exp\{-\frac{1}{2}\mathrm{trace}\{\boldsymbol{R}\boldsymbol{\Sigma}^{-1}\}\}, \tag{4.1}$$

where

$$c_0^{-1} = 2^{(k-r-1)r/2}\pi^{r(r-1)/4}\prod_{j=1}^{r}\Gamma\Big(\frac{k-r-j}{2}\Big),$$

and $2r < k$. So

$$\int_{\Omega}|\boldsymbol{\Sigma}|^{-k/2}\exp\{-\frac{1}{2}\mathrm{trace}\{\boldsymbol{R}\boldsymbol{\Sigma}^{-1}\}\}\,d\boldsymbol{\Sigma} = c_0^{-1}|\boldsymbol{R}|^{-(k-r-1)/2}, \tag{4.2}$$

where $\Omega = \{\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r} : \boldsymbol{\Sigma} > \mathbf{0}\}$. The notation employed is $\boldsymbol{\Sigma} \sim \mathrm{W}_r^{-1}[k, \boldsymbol{R}]$.

Equation (4.1) can take the form

$$p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(r+n/2)}\exp\{-\frac{1}{2}\mathrm{trace}\{n\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\}\}, \tag{4.3}$$

where $n = k - 2r$, $\boldsymbol{S} = n^{-1}\boldsymbol{R}$ and defines the inverse Wishart distribution with n degrees of freedom. Both equations (4.1) and (4.3) express the same distribution, which in the latter case will be denoted by $\mathrm{W}_n^{-1}[\boldsymbol{S}]$. In [34] and [51, chapter 16] it is stated that

$$\mathrm{E}[\boldsymbol{\Sigma}^{-1}] = \boldsymbol{S}^{-1}.$$

We will show that this is not true in the multivariate case, but it holds only in the limiting form. Assume that $\Sigma \sim \mathrm{W}_n^{-1}[\boldsymbol{S}]$ or $\Sigma \sim \mathrm{W}_r^{-1}[k, \boldsymbol{R}]$, with $\boldsymbol{R}$, k as defined before. Then, it is well known (see [32, chapter 5]) that $\Sigma^{-1} \sim \mathrm{W}_r[k - r - 1, \boldsymbol{R}^{-1}]$ (the Wishart distribution with $k - r - 1$ degrees of freedom and parameter matrix $\boldsymbol{R}^{-1}$, see page 66) and so the expectation of Σ^{-1} is $\mathrm{E}[\Sigma^{-1}] = (k - r - 1)\boldsymbol{R}^{-1}$. This proves that

$$\mathrm{E}[\Sigma^{-1}] = \frac{n + r - 1}{n}\boldsymbol{S}^{-1}.$$

Note that when $r = 1$, $\mathrm{E}[\Sigma^{-1}] = \boldsymbol{S}^{-1}$, and when $r \geq 2$, $\lim_{n\to\infty} \mathrm{E}[\Sigma^{-1}] = \boldsymbol{S}^{-1}$. This is illustrated in Figure 4.1, where the real sequence $\{x_n\}$, with

$$x_n = \frac{n + r - 1}{n}a,$$

is drawn for three values of r, ($r = 10, 20, 50$, as is shown from the bottom) and $a = 1$. These three lines correspond to $\mathrm{E}[\sigma^{ii}] = x_n$, where $\Sigma^{-1} = \{\sigma^{ij}\}$, $(i, j = 1, \ldots, r)$.

Figure 4.1 indicates that even if the dimensionality of Σ is small, say $r = 10$, for relatively low degrees of freedom, say $n = 10$, the expectation of Σ^{-1} is almost twice as large as stated in [51, chapter 16].

Lemma 4.1. *Consider an $r \times r$ SPD (symmetric positive definite) random matrix Σ, an $r \times r$ SPD matrix $\boldsymbol{S}$, and an $r \times r$ diagonal matrix $\boldsymbol{N}$ with positive diagonal elements. Then the function*

$$p(\Sigma) \propto |\Sigma|^{-\left(r + \frac{trace(\boldsymbol{N})}{2r}\right)} exp\{-\frac{1}{2} trace\{\boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2}\Sigma^{-1}\}\} \tag{4.4}$$

defines a density function and hence a distribution.

Proof. Using the bijective transformation

$$\boldsymbol{R} = \boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2} \qquad \text{and} \qquad k = 2r + \frac{\mathrm{trace}(\boldsymbol{N})}{r}, \tag{4.5}$$

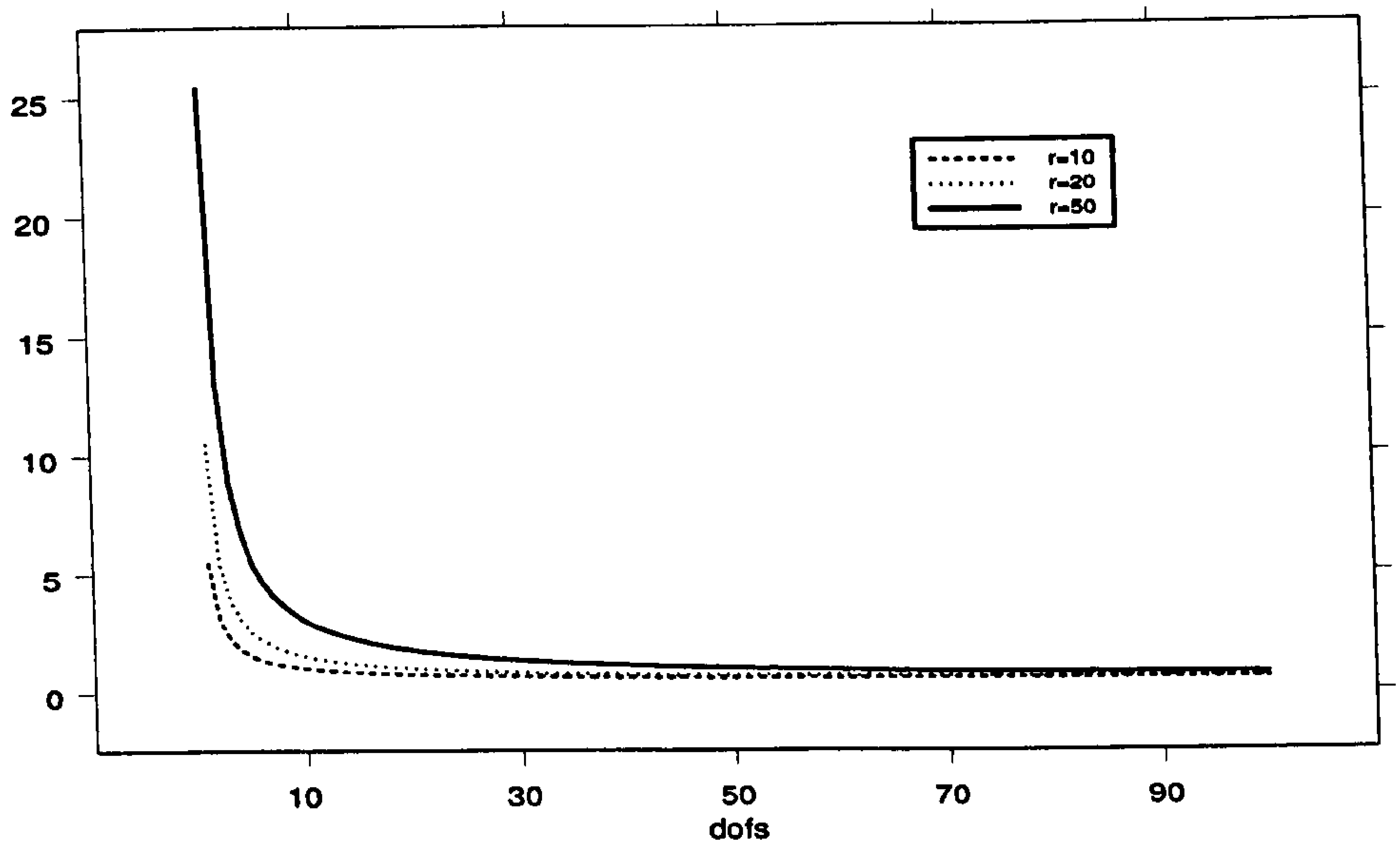

Figure 4.1: Limiting behaviour of the expectation of an inverse Wishart random matrix

(4.4) is directly obtained from (4.1). $\square$

Note that if all diagonal elements of $\boldsymbol{N}$ are the same and equal to say n, then equation (4.4) reduces to (4.3). The normalizing constant of (4.4) is

$$c_1 = c_0 |\boldsymbol{S}|^{\left(r + \frac{\text{trace}(\boldsymbol{N})}{r} - 1\right)/2} \left(\prod_{j=1}^{r} n_j\right)^{\left(r + \frac{\text{trace}(\boldsymbol{N})}{r} - 1\right)/2},$$

where

$$c_0^{-1} = 2^{\left(r + \frac{\text{trace}(\boldsymbol{N})}{r} - 1\right) r/2} \pi^{r(r-1)/4} \prod_{j=1}^{r} \Gamma\left(\frac{r + \frac{\text{trace}(\boldsymbol{N})}{r} - j}{2}\right),$$

and $\boldsymbol{N} = \text{diag}\{n_1, \ldots, n_r\}$, $(n_i > 0; i = 1, \ldots, r)$.

Definition 4.1. *An $r \times r$ SPD random matrix Σ is said to follow the* **generalized inverse Wishart** *distribution with a matrix $\boldsymbol{N}$ of degrees of* *freedom and parameter an SPD matrix S if and only if its density is given by equation (4.4).*

This will be written employing the notation $\Sigma \sim \mathrm{GW}^{-1}[S, N, m]$, where

$$m = r + \frac{\operatorname{trace}(N)}{2r}. \tag{4.6}$$

The first and second moments of Σ can be derived by the standard results ([32, chapter 5]) and transformation (4.5) as follows.

The mean of Σ is

$$\mathrm{E}[\Sigma] = \frac{1}{k-2r-2} R = \left(\frac{\operatorname{trace}(N)}{r} - 2\right)^{-1} N^{1/2} S N^{1/2},$$

for $r^{-1}\operatorname{trace}(N) > 2$.

Now for the variance write $\Sigma = \{\sigma_{ij}\}$, $S = \{s_{ij}\}$, $R = \{R_{ij}\}$, $i, j = 1, \ldots, r$. Then the covariances are

$$\begin{aligned} \mathrm{C}[\sigma_{ij}, \sigma_{kl}] &= \frac{2(k-2r-2)^{-1} R_{ij} R_{kl} + R_{ik} R_{jl} + R_{il} R_{kj}}{(k-2r-1)(k-2r-2)(k-2r-4)} \\ &= 2\sqrt{n_i n_j n_k n_l}\left(\frac{\operatorname{trace}(N)}{r} - 1\right)^{-1}\left(\frac{\operatorname{trace}(N)}{r} - 2\right)^{-1} \\ &\quad \times \left(\frac{\operatorname{trace}(N)}{r} - 4\right)^{-1}\left[\left(\frac{\operatorname{trace}(N)}{r} - 2\right)^{-1} s_{ij} s_{kl} + s_{ik} s_{jl} \right. \\ &\quad \left. + s_{il} s_{kj}\right], \end{aligned}$$

for $r^{-1}\operatorname{trace}(N) > 4$, $(k - 2r - 4 > 0)$.

Assume now the following partition of Σ, S, and N.

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}, \quad S = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{pmatrix}, \quad N = \begin{pmatrix} N_{11} & \mathrm{O} \\ \mathrm{O}' & N_{22} \end{pmatrix},$$

where $\dim(\Sigma_{11}) = \dim(S_{11}) = \dim(N_{11}) = q \times q$, $\dim(\Sigma_{12}) = \dim(S_{12}) = q \times (r-q)$, $\dim(\Sigma_{22}) = \dim\,(S_{22}) = \dim(N_{22}) = (r-q) \times (r-q)$, for some $1 \leq q \leq r-1$.

Theorem 4.1. *If the SPD matrix Σ follows a generalized inverse Wishart distribution with a matrix of degrees of freedom $\boldsymbol{N}$ and parameter matrix $\boldsymbol{S}$ under the above partition of Σ, $\boldsymbol{S}$, $\boldsymbol{N}$, the distribution of Σ_{11} is $\Sigma_{11} \sim GW^{-1}[\boldsymbol{S}_{11}, \boldsymbol{N}_{11}, m_1]$, with $m_1 = q + \frac{trace(\boldsymbol{N})}{2r}$.*

Proof. The proof suggests the use of the transformation (4.5) together with the partition of $\boldsymbol{R}$ of equation (4.1) as

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{12}' & \boldsymbol{R}_{22} \end{pmatrix},$$

where $\dim(\boldsymbol{R}_{11}) = q \times q$, $\dim(\boldsymbol{R}_{12}) = q \times (r-q)$, $\dim(\boldsymbol{R}_{22}) = (r-q) \times (r-q)$. Now using the known marginal distributional results of the inverse Wishart distribution $\mathrm{W}_r^{-1}[k, \boldsymbol{R}]$ (see [14, chapter 3]) and upon noticing

$$\boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2} = \begin{pmatrix} \boldsymbol{N}_{11}^{1/2}\boldsymbol{S}_{11}\boldsymbol{N}_{11}^{1/2} & \boldsymbol{N}_{11}^{1/2}\boldsymbol{S}_{12}\boldsymbol{N}_{22}^{1/2} \\ \boldsymbol{N}_{22}^{1/2}\boldsymbol{S}_{12}'\boldsymbol{N}_{11}^{1/2} & \boldsymbol{N}_{22}^{1/2}\boldsymbol{S}_{22}\boldsymbol{N}_{22}^{1/2} \end{pmatrix}, \tag{4.7}$$

it is easily deduced that $\Sigma_{11} \sim \mathrm{GW}^{-1}[\boldsymbol{S}_{11}, \boldsymbol{N}_{11}, m_1]$, where $m_1 = m + q - r = q + \frac{\mathrm{trace}(\boldsymbol{N})}{2r}$. □

Theorem 4.2. *If Θ is an $n \times r$ random matrix that follows a matrix normal distribution, $(\Theta|\Sigma) \sim N[\boldsymbol{m}, \boldsymbol{C}, \Sigma]$, where $\boldsymbol{m}$ is an $n \times r$ matrix, $\boldsymbol{C}$ an $n \times n$ left variance matrix, Σ an $r \times r$ right variance matrix such that $\Sigma \sim GW^{-1}[\boldsymbol{S}, \boldsymbol{N}, m]$, for a non-singular $r \times r$ matrix $\boldsymbol{S}$, $\boldsymbol{N}$ a diagonal matrix with positive diagonal elements, and m as defined in (4.6), then the posterior distribution of Σ given Θ is*

$$(\Sigma|\Theta) \sim GW^{-1}[\boldsymbol{S}^*, \boldsymbol{N}^*, m^*],$$

where $\boldsymbol{N}^* = \boldsymbol{N} + n\boldsymbol{I}$, $\boldsymbol{N}^{*1/2}\boldsymbol{S}^*\boldsymbol{N}^{*1/2} = (\Theta - \boldsymbol{m})'\boldsymbol{C}^{-1}(\Theta - \boldsymbol{m}) + \boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2}$, and $m^* = r + \frac{trace(\boldsymbol{N}^*)}{2r}$.

Proof. Form the joint distribution of Θ and Σ and write

$$\begin{aligned}
p(\Sigma|\Theta) &\propto p(\Theta,\Sigma) \\
&\propto |\Sigma|^{-\left(r+\frac{n}{2}+\frac{\text{trace}(N)}{2r}\right)}\exp\{-\frac{1}{2}\text{trace}\{[(\Theta-m)'C^{-1}(\Theta-m) \\
&\quad +N^{1/2}SN^{1/2}]\Sigma^{-1}\}\},
\end{aligned}$$

which is sufficient for the proof under the definitions of N^*, S^*, and m^*. □

Now we are interested in the distribution of the inverse matrix of Σ. Set $V=\Sigma^{-1}$ and see that $|\Sigma|=|V|^{-1}$. Verify from formula (A.4) that

$$\left|\frac{\partial \text{vech}(V^{-1})}{\partial(\text{vech}V)'}\right| = (-1)^{r(r+1)/2}|V|^{-(r+1)}.$$

So

$$p(V)\propto |V|^{\frac{\text{trace}(N)}{2r}-1}\exp\{-\frac{1}{2}\text{trace}\{N^{1/2}SN^{1/2}V\}\}. \tag{4.8}$$

It can be shown that equation (4.8) defines a distribution. The equivalence with the standard Wishart distribution is as follows. According to [32] or [14], the non-singular Wishart distribution is expressible as

$$p(V)=c|R|^{-k/2}|V|^{(k-r-1)/2}\exp\{-\frac{1}{2}\text{trace}\{R^{-1}V\}\}, \tag{4.9}$$

where V is an $r\times r$ SPD random matrix, R an SPD parameter matrix, and

$$c=\left[2^{kr/2}\pi^{r(r-1)/4}\prod_{j=1}^{r}\Gamma\left(\frac{k+1-j}{2}\right)\right]^{-1},$$

with $r\leq k$. By means of notation, $V\sim \text{W}_r[k,R]$, meaning that V has a Wishart distribution with k degrees of freedom and parameter matrix R.

Similarly as for the inverse Wishart distribution we can set $k=r+n-1$ and $R^{-1}=nS$. Then equation (4.9) becomes

$$p(V)\propto |V|^{n/2-1}\exp\{-\frac{1}{2}\text{trace}\{nSV\}\}, \tag{4.10}$$

where $n \geq 1$.

Equations (4.9), (4.10) express the same distribution, which using the latter notation is denoted by $\mathrm{W}_n[\boldsymbol{S}]$ and is referred to as the Wishart distribution with n degrees of freedom and parameter matrix $\boldsymbol{S}$.

By using the bijective transformation $k = r + \frac{\mathrm{trace}(\boldsymbol{N})}{r} - 1$ and $\boldsymbol{R}^{-1} = \boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2}$, equation (4.8) is obtained by (4.9). Now the definition comes naturally.

Definition 4.2. *The $r \times r$ SPD matrix $\boldsymbol{V}$ is said to follow the* **generalized Wishart** *distribution with a matrix $\boldsymbol{N}$ of degrees of freedom and parameter matrix $\boldsymbol{S}$ if and only if its density is given by equation (4.8).*

Note that when $r = 1$ we obtain the gamma distribution $\mathrm{G}[\frac{n_1}{2}, \frac{n_1 S}{2}]$ and when $n_1 = \cdots = n_r = n$ we get the Wishart distribution with n degrees of freedom $\mathrm{W}_n[\boldsymbol{S}]$. The notation is $\boldsymbol{V} \sim \mathrm{GW}[\boldsymbol{S}, \boldsymbol{N}, m]$, where $m = \frac{\mathrm{trace}(\boldsymbol{N})}{2r} - 1$. From the above discussion it is clear that if $\boldsymbol{\Sigma} \sim \mathrm{GW}^{-1}[\boldsymbol{S}, \boldsymbol{N}, m]$, then $\boldsymbol{\Sigma}^{-1} \sim \mathrm{GW}[\boldsymbol{S}, \boldsymbol{N}, m - r - 1]$. Similarly, it can be shown that if $\boldsymbol{\Sigma} \sim \mathrm{GW}[\boldsymbol{S}, \boldsymbol{N}, m]$, then $\boldsymbol{\Sigma}^{-1} \sim \mathrm{GW}^{-1}[\boldsymbol{S}, \boldsymbol{N}, m + r + 1]$.

Let $\boldsymbol{\Sigma} \sim \mathrm{GW}^{-1}[\boldsymbol{S}, \boldsymbol{N}, m]$, with $\boldsymbol{S}$, $\boldsymbol{N}$, m as defined before. In terms of the standard Wishart distribution, this may be written as
$\boldsymbol{\Sigma}^{-1} \sim \mathrm{W}_r[k - r - 1, \boldsymbol{R}^{-1}]$, with $k, \boldsymbol{R}$ as in equation (4.5). Thus

$$\mathrm{E}[\boldsymbol{\Sigma}^{-1}] = \Big(\frac{\mathrm{trace}\,(\boldsymbol{N})}{r} + r - 1\Big)\boldsymbol{N}^{-1/2}\boldsymbol{S}^{-1}\boldsymbol{N}^{-1/2}.$$

In parallel with Appendix A.5, if $\boldsymbol{l}'\boldsymbol{N}\boldsymbol{l} \to \infty$ for all $\boldsymbol{l} \in \mathbb{R}^r$, then for at least one of n_i $(i = 1, \ldots, r)$, $n_i \to \infty$. We will use the notation $\boldsymbol{N} \to \infty$ to signify that all $n_i \to \infty$, $(i = 1, \ldots, r)$.

Then, the limit of a matrix-valued function with respect to $\boldsymbol{N}$, is a direct extension to several variables (diagonal elements of $\boldsymbol{N}$) of Definition A.3.

The following theorem of the section gives the limit of $\mathrm{E}[\Sigma^{-1}]$.

Theorem 4.3. *Let Σ be an $r \times r$ SPD matrix with a generalized inverse Wishart distribution, $\Sigma \sim GW^{-1}[S, N, m]$, where $N = diag\{n_1, \ldots, n_r\}$, for any positive real numbers n_i, $(i = 1, \ldots, r)$. Let all n_i have the same rate of convergence so that as, $N \to \infty$, $\max\{n_i\}/\min\{n_i\} \to 1$. Then as $N \to \infty$*

$$\Big(\frac{trace\,(N)}{r} + r - 1\Big) N^{-1/2} S^{-1} N^{-1/2} \to S^{-1}.$$

Proof. Write $S^{-1} = \{s^{ij}\}$, $(1 \le i, j \le r)$ and

$$X_N = \left(\frac{\mathrm{trace}(N)}{r} + r - 1\right) N^{-1/2} S^{-1} N^{-1/2}.$$

According to the above discussion and Definition A.5 it remains to prove

$$l' X_N l \to l' S^{-1} l, \tag{4.11}$$

for any vector $l \in \mathbb{R}^r$.

Let $m = \max\{n_1, \ldots, n_r\}$ and $k = \min\{n_1, \ldots, n_r\}$, so that $m, k \to \infty$. Then

$$l' X_N l \le \sum_{i=1}^{r} \sum_{j=1}^{r} l_i l_j s^{ij} \frac{m + r - 1}{\sqrt{n_i n_j}},$$

and

$$l' X_N l \ge \sum_{i=1}^{r} \sum_{j=1}^{r} l_i l_j s^{ij} \frac{k + r - 1}{\sqrt{n_i n_j}},$$

The hypothesis of the theorem implies that both $(m+r-1)/\sqrt{n_i n_j} \to 1$ and $(k + r - 1)/\sqrt{n_i n_j} \to 1$. This with the above inequalities proves equation (4.11). □

In this section generalizations of the Wishart and inverse Wishart distributions are proposed. These will provide a comprehensive missing observation analysis and variance intervention in later chapters. Here, a brief discussion of other generalizations are done and it is explained why the proposed distributions are not obtained by these older generalizations.

Two of the latest and most used extensions of the inverse Wishart distribution will be discussed. The first is the so called *hyper inverse Wishart* distribution. This distribution was originally defined in [9] and further applied in [38]. The development of this extension is within the graphical model arena, and so the mathematical details are not included in this thesis. The important note is that this generalization concerns only the parameter SPD matrix of the usual inverse Wishart distribution, while the degrees of freedom remain scalar. This clearly shows that the generalized inverted Wishart distribution, proposed in this thesis, is not embedded in the above family of distributions.

The second extension of the inverse Wishart distribution appeared initially in [6, chapter 7] and further applied in [13]. Let $\Sigma \sim \mathrm{W}_r^{-1}[k, \boldsymbol{R}]$ with $k, \boldsymbol{R}$ as in page 61. Dawid ([8]) proposes that we shall use the notation $\Sigma \sim \mathrm{IW}(\delta; \boldsymbol{R})$ instead of the usual $\Sigma \sim \mathrm{W}_r^{-1}[k, \boldsymbol{R}]$, with $\delta = k - 2r$. The factor δ is no more the degrees of freedom, but a hyperparameter. Then the following result applies, as it appears in [8].

Lemma 4.2. *Let $\Sigma \sim IW(\delta; \boldsymbol{G})$, for some known $\boldsymbol{G}, \delta$, and Σ is partitioned as $\Sigma = \{\Sigma_{ij}\}$, for $i, j = 1, 2$. Define $\boldsymbol{B} = \Sigma_{22}^{-1}\Sigma_{21}$ and $\Gamma = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Then the following apply*

(a) Σ_{22} is independent of $(\boldsymbol{B}, \Gamma)$ and $\Sigma_{22} \sim IW(\delta; \boldsymbol{G}_{22})$;

(b) $\Gamma \sim IW(\delta + r - q; \boldsymbol{G}_{11} - \boldsymbol{G}_{12}\boldsymbol{G}_{22}^{-1}\boldsymbol{G}_{21})$;

(c) $(\boldsymbol{B}|\Gamma) \sim N[\boldsymbol{G}_{22}^{-1}\boldsymbol{G}_{21}, \boldsymbol{G}_{22}^{-1}, \Gamma]$,

where $\boldsymbol{G} = \{\boldsymbol{G}_{ij}\}$, $(i, j = 1, 2)$, *and* $dim(\Sigma_{11}) = q$, $dim(\Sigma_{22}) = (r-q) \times (r-q)$.

Proof. See [8]. $\square$

Let $\boldsymbol{Y}$ be an $r \times m$ observation matrix. A matrix normal regression model is used, given by

$$(\boldsymbol{Y}'|\Theta, \Sigma) \sim \mathrm{N}[\boldsymbol{F}'\Theta, \boldsymbol{V}, \Sigma],$$

where the matrices $\boldsymbol{F}$, $\boldsymbol{V}$ are known and Θ is a matrix of states.

Considering the same partition of Σ and $\boldsymbol{G}$ as in the above lemma, the *generalized inverse Wishart* distribution with 2 blocks $(k = 2)$ is defined to be the distribution with the following structure

$$\Sigma_{22} \perp\!\!\!\perp (\boldsymbol{B}, \boldsymbol{\Gamma}),$$

$$\Sigma_{22} \sim \mathrm{IW}(\delta_2; \boldsymbol{G}_{22}),$$

$$\boldsymbol{\Gamma} \sim \mathrm{IW}(\delta_1 + r - q; \boldsymbol{Q}),$$

$$(\boldsymbol{B}|\boldsymbol{\Gamma}) \sim \mathrm{N}[\boldsymbol{B}_0, \boldsymbol{H}, \boldsymbol{\Gamma}],$$

for some extra parameters $\boldsymbol{Q}, \boldsymbol{B}_0, \boldsymbol{H}$, and hyperparameters $\delta_1, \delta_2 > 0$.

It is clear that if $\delta_1 = \delta_2$, $\boldsymbol{Q} = \boldsymbol{G}_{11} - \boldsymbol{G}_{12}\boldsymbol{G}_{22}^{-1}\boldsymbol{G}_{21}$, $\boldsymbol{B}_0 = \boldsymbol{G}_{22}^{-1}$, and $\boldsymbol{H} = \boldsymbol{G}_{22}^{-1}\boldsymbol{G}_{21}$, then this distribution is reduced to an inverse Wishart distribution.

The general case of this distribution for any $k \geq 2$ blocks is defined analogously with the $k = 2$ case, just by defining random quantities $\Sigma_{(ii)}$, $(i = 1, \ldots, k)$, as

$$\Sigma_{(ii)} = \begin{pmatrix} \Sigma_{ii} & \Sigma_{i(i+1)} \\ \Sigma_{(i+1)i} & \Sigma_{([i+1][i+1])} \end{pmatrix},$$

where $\Sigma_{i(i+1)}$ is comfortably defined. $\Sigma_{(ii)}$ is the right variance matrix of $\boldsymbol{Y}_{(i)} = (\boldsymbol{Y}_i, \ldots, \boldsymbol{Y}_k) = (\boldsymbol{Y}_i, \boldsymbol{Y}_{(i+1)})$, where $\boldsymbol{Y}_i$ is the $r \times q_i$ matrix such that $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_k)$, $(i = 1, \ldots, k)$ and $q_1 + \cdots + q_k = m$.

Define $\boldsymbol{B}_i = (\Sigma_{([i+1][i+1])})^{-1}\Sigma_{(i+1)i}$, $\boldsymbol{\Gamma}_i = \Sigma_{ii} - \Sigma_{i(i+1)}(\Sigma_{([i+1][i+1])})^{-1}\Sigma_{(i+1)i}$, for $i = 1, \ldots, k-1$. Then Σ is said to follow a generalized inverse Wishart distribution (GIW) with $k \geq 2$ blocks, if and only if Σ_{kk} is independent of $(\boldsymbol{B}_i, \boldsymbol{\Gamma}_i)$, where the later are independent of each other and

$$\Sigma_{kk} \sim \mathrm{IW}(\delta_k; \boldsymbol{G}_{kk}),$$

$$\boldsymbol{\Gamma}_i \sim \mathrm{IW}(\delta_i + q_{(i)}; \boldsymbol{Q}_i),$$

$$(\boldsymbol{B}_i | \boldsymbol{\Gamma}_i) \sim \mathrm{N}[\boldsymbol{B}_{0,i}, \boldsymbol{H}_i, \boldsymbol{\Gamma}_i],$$

for some hyperparameters $\delta_k, \delta_i, \boldsymbol{G}_{kk}, \boldsymbol{Q}_i, \boldsymbol{B}_{0,i}, \boldsymbol{H}_i$, $(i = 1, \ldots, k-1)$.

It follows that with the partition of Σ

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix}$$

the distribution of Σ_{ii} $(i = 1, 2)$ is an inverse Wishart distribution, if a generalized inverse Wishart distribution for Σ is assumed. If, on the other hand Σ follows a generalized inverse Wishart distribution (as defined in this chapter) the diagonal elements Σ_{ii} follow again general inverse Wishart distributions and not inverse Wisharts (see Theorem 4.1).

There are some further problems related with the above extensions (hyper and generalized inverse Wishart distributions). Although generalizations have been proposed for the inverse Wishart, it is not clear whether these analyses lead to generalizations to related distributions with the normal and Wishart, e.g. the matrix T distribution.

This point is the linking one with the next section. We show that using the proposed generalized inverse Wishart distribution a new class of matrix T distributions can be defined. And these generalizations are very natural providing all the details for the specification of the parameters of the various distributions. This is far to be an easy task for the hyper inverse Wishart and the generalized inverse Wishart distributions, see [9, 13]. It seems that the incorporation of a matrix of degrees of freedom in the new distributions gives much versatility keeping at the same time the conjugacy and many of the identities and characterizations of the usual Wishart and related distributions. It is believed that the new generalizations will provide a more general and widely applicable toolkit that is why the name "generalized" has been used.

4.3 Generalized T Distribution

In a similar fashion to that in Section 4.2, the matrix T distribution can be extended to incorporate a matrix of degrees of freedom.

Definition 4.3. *An $n \times r$ random matrix Θ is said to follow the* **generalized T** *distribution with mode matrix $\boldsymbol{m}$, left scale matrix $\boldsymbol{C}$, right scale matrix $\boldsymbol{S}$, and a matrix of degrees of freedom $\boldsymbol{N}$ if and only if its density is expressed by*

$$p(\Theta) \propto |\boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2} + (\Theta - \boldsymbol{m})'\boldsymbol{C}^{-1}(\Theta - \boldsymbol{m})|^{-(r+\frac{trace(\boldsymbol{N})}{r}+n-1)/2}. \quad (4.12)$$

This distribution can be derived from the matrix T distribution (see equation (B.3)) by setting $\boldsymbol{Q} = \boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2}$, $\boldsymbol{T} = \Theta$, $\boldsymbol{M} = \boldsymbol{m}$, $\boldsymbol{P} = \boldsymbol{C}$, and

$k = \frac{\text{trace}(\boldsymbol{N})}{r}$. But also (B.3) is obtainable by (4.12). The normalizing constant of (4.12) is obtainable from (B.3) as

$$c_2 = \frac{|\boldsymbol{S}|^{(p+r-1)/2}\left(\prod_{j=1}^{r} n_j\right)^{(p+r-1)/2}|\boldsymbol{C}|^{-r/2}}{c(p,n,r)},$$

with $p = \frac{\text{trace}(\boldsymbol{N})}{r}$, where $c(p,n,r)$ is defined as in (B.4). Note that if all the diagonal elements of $\boldsymbol{N}$ are the same, the distribution is a matrix T distribution. The adopted notation is $\Theta \sim \text{GT}[\boldsymbol{m}, \boldsymbol{C}, \boldsymbol{S}, \boldsymbol{N}, p]$, with p as defined before.

Theorem 4.4. *Let Θ be an $n \times r$ random matrix that follows a matrix normal distribution conditional on Σ, and Σ an $r \times r$ SPD random matrix that follows a generalized inverse Wishart distribution, written $(\Theta|\Sigma) \sim N[\boldsymbol{m}, \boldsymbol{C}, \Sigma]$ and $\Sigma \sim GW^{-1}[\boldsymbol{S}, \boldsymbol{N}, m]$ respectively, for known quantities $\boldsymbol{m}$, $\boldsymbol{C}, \boldsymbol{N}, \boldsymbol{S}, m$. Then, the marginal distribution of Θ is a $GT[\boldsymbol{m}, \boldsymbol{C}, \boldsymbol{S}, \boldsymbol{N}, p]$.*

Proof. Write down the joint distribution of Θ and Σ

$$\begin{aligned} p(\Theta, \Sigma) &= p(\Theta|\Sigma)p(\Sigma) \\ &\propto |\Sigma|^{-\left(r+\frac{n}{2}+\frac{\text{trace}(\boldsymbol{N})}{2r}\right)}\exp\{-\frac{1}{2}\text{trace}\{[(\Theta-\boldsymbol{m})'\boldsymbol{C}^{-1}(\Theta-\boldsymbol{m}) \\ &\quad +\boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2}]\Sigma^{-1}\}\}. \end{aligned}$$

Now the marginal distribution of Θ will be

$$p(\Theta) = \int_{\Omega} p(\Theta, \Sigma)\, d\Sigma,$$

where $\Omega = \{\Sigma \in \mathbb{R}^{r \times r} : \Sigma > O\}$.

Set $\boldsymbol{R} = (\Theta - \boldsymbol{m})'\boldsymbol{C}^{-1}(\Theta - \boldsymbol{m}) + \boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2}$ and $k = 2r + n + \frac{\text{trace}(\boldsymbol{N})}{r}$ and from equation (4.2) we obtain equation (4.12). □

The marginal and conditional distributions of the generalized T are derived:

Let $\Theta \sim \mathrm{GT}[\boldsymbol{m}, \boldsymbol{C}, \boldsymbol{S}, \boldsymbol{N}, p]$, with parameters as defined above. Also, partition Θ, $\boldsymbol{m}$, $\boldsymbol{N}$, and $\boldsymbol{S}$, as follows.

$$\Theta = (\Theta_1, \Theta_2), \qquad \boldsymbol{m} = (\boldsymbol{m}_1, \boldsymbol{m}_2),$$

$$\boldsymbol{N} = \begin{pmatrix} \boldsymbol{N}_{11} & \mathrm{O} \\ \mathrm{O}' & \boldsymbol{N}_{22} \end{pmatrix}, \qquad \boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}'_{12} & \boldsymbol{S}_{22} \end{pmatrix},$$

where $\dim(\Theta_1) = \dim(\boldsymbol{m}_1) = n \times q$, $\dim(\Theta_2) = \dim(\boldsymbol{m}_2) = n \times (r-q)$, $\dim(\boldsymbol{N}_{11}) = \dim(\boldsymbol{S}_{11}) = q \times q$, $\dim(\boldsymbol{S}_{12}) = q \times (r-q)$, and $\dim(\boldsymbol{N}_{22}) = \dim(\boldsymbol{S}_{22}) = (r-q) \times (r-q)$, for some $1 \le q \le r-1$.

Using the correspondence between the general T and the matrix T distributions and equations (B.5), (B.6) together with equation (4.7) we derive the marginal and conditional distributions of the general T distribution as follows.

The marginal distribution of Θ_2 is

$$\Theta_2 \sim \mathrm{GT}[\boldsymbol{m}_2, \boldsymbol{C}, \boldsymbol{S}_{22}, \boldsymbol{N}_{22}, p] \tag{4.13}$$

and the conditional distribution of Θ_2, given Θ_1 is

$$(\Theta_2 | \Theta_1) \sim \mathrm{GT}[\boldsymbol{m}_{2|1}, \boldsymbol{C}_2, \boldsymbol{S}_{2|1}, \boldsymbol{N}_{22}, p_2], \tag{4.14}$$

where

$$\boldsymbol{m}_{2|1} = \boldsymbol{m}_2 + (\Theta_1 - \boldsymbol{m}_1)\boldsymbol{N}_{11}^{-1/2}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}\boldsymbol{N}_{22}^{1/2},$$

$$\boldsymbol{C}_2 = \boldsymbol{C} + (\Theta_1 - \boldsymbol{m}_1)\boldsymbol{N}_{11}^{-1/2}\boldsymbol{S}_{11}^{-1}\boldsymbol{N}_{11}^{-1/2}(\Theta_1 - \boldsymbol{m}_1)',$$

$$\boldsymbol{S}_{22|1} = \boldsymbol{S}_{22} - \boldsymbol{S}'_{12}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12},$$

and $p_2 = p + q$, provided that $\boldsymbol{S}_{11}$ is non-singular.

4.4 Matrix Normal DLMs

The common components model is the only multivariate Normal DLM with observational variances that allows a conjugate analysis, as stated in [34] and [4]. This model was defined and briefly discussed in Section 2.3.2.

The analysis of this model uses the inverse Wishart distribution, hence it carries all its limitations (for a discussion of these limitations see Section 6.2). Here, we extend the model with the introduction of the generalized inverse Wishart distribution and its relevant distributions as developed in Sections 4.2 and 4.3. So the current model is

$$\boldsymbol{Y}'_t = \boldsymbol{F}'_t\Theta_t + \boldsymbol{\nu}'_t, \qquad \boldsymbol{\nu}'_t \sim \mathrm{N}[\mathbf{0}, V_t, \Sigma], \tag{4.15}$$

$$\Theta_t = \boldsymbol{G}_t\Theta_{t-1} + \boldsymbol{\Omega}_t, \qquad \boldsymbol{\Omega}_t \sim \mathrm{N}[\mathbf{O}, \boldsymbol{W}_t, \Sigma], \tag{4.15$'$}$$

and

$$(\Theta_0, \Sigma|D_0) \sim \mathrm{NGW}^{-1}[\boldsymbol{m}_0, \boldsymbol{C}_0, \boldsymbol{S}_0, \boldsymbol{N}_0, m_0], \tag{4.16}$$

for some known defining parameters $\boldsymbol{m}_0$, $\boldsymbol{C}_0$, $\boldsymbol{S}_0$ and $\boldsymbol{N}_0 = \mathrm{diag}\{n_{10}, \dots, n_{r0}\}$. The "NGW^{-1}" (normal generalized inverse Wishart) distribution is discussed in Section B.2. This model is referred to as ECCM (Extended Common Components Model). Then, for all times $t \geq 1$, the following results apply.

Theorem 4.5. *One-step forecast and posterior distributions in the model (4.15), (4.15$'$), and (4.16) are given, for each t, as follows.*

(a) Posterior at $t-1$:

For some $\boldsymbol{m}_{t-1}$, $\boldsymbol{C}_{t-1}$, $\boldsymbol{S}_{t-1}$, and $\boldsymbol{N}_{t-1}$,

$$(\Theta_{t-1}, \Sigma|D_{t-1}) \sim NGW^{-1}[\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}, \boldsymbol{S}_{t-1}, \boldsymbol{N}_{t-1}, m_{t-1}],$$

where $m_{t-1} = r + \frac{trace(\boldsymbol{N}_{t-1})}{2r}$.

(b) Prior at t:

$$(\Theta_t, \Sigma | D_{t-1}) \sim NGW^{-1}[\boldsymbol{a}_t, \boldsymbol{R}_t, \boldsymbol{S}_{t-1}, \boldsymbol{N}_{t-1}, m_{t-1}],$$

where

$$\boldsymbol{a}_t = \boldsymbol{G}_t \boldsymbol{m}_{t-1} \qquad \text{and} \qquad \boldsymbol{R}_t = \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t' + \boldsymbol{W}_t.$$

(c) One-step forecast:

$$(\boldsymbol{Y}_t' | \Sigma, D_{t-1}) \sim N[\boldsymbol{f}_t', Q_t, \Sigma],$$

with marginal

$$(\boldsymbol{Y}_t' | D_{t-1}) \sim GT[\boldsymbol{f}_t', Q_t, \boldsymbol{S}_{t-1}, \boldsymbol{N}_{t-1}, p_{t-1}],$$

where

$$\boldsymbol{f}_t' = \boldsymbol{F}_t' \boldsymbol{a}_t, \qquad Q_t = \boldsymbol{F}_t' \boldsymbol{R}_t \boldsymbol{F}_t + V_t, \qquad \text{and} \qquad p_{t-1} = \frac{trace(\boldsymbol{N}_{t-1})}{r}.$$

(d) Posterior at t:

$$(\Theta_t, \Sigma | D_t) \sim NGW^{-1}[\boldsymbol{m}_t, \boldsymbol{C}_t, \boldsymbol{S}_t, \boldsymbol{N}_t, m_t],$$

with

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t \boldsymbol{e}_t', \qquad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t Q_t \boldsymbol{A}_t',$$

$$\boldsymbol{N}_t = \boldsymbol{N}_{t-1} + \boldsymbol{I}, \qquad \boldsymbol{N}_t^{1/2} \boldsymbol{S}_t \boldsymbol{N}_t^{1/2} = \boldsymbol{N}_{t-1}^{1/2} \boldsymbol{S}_{t-1} \boldsymbol{N}_{t-1}^{1/2} + \boldsymbol{e}_t Q_t^{-1} \boldsymbol{e}_t',$$

$$m_t = r + \frac{trace(\boldsymbol{N}_t)}{2r},$$

where

$$\boldsymbol{A}_t = \boldsymbol{R}_t \boldsymbol{F}_t Q_t^{-1} \qquad \text{and} \qquad \boldsymbol{e}_t = \boldsymbol{Y}_t - \boldsymbol{f}_t.$$

Proof. The proof is by induction. Suppose that (a) is true. (b) follows immediately from (a) and the evolution equation. For (c) first see that (b) and the observation equation lead to $\mathrm{E}[\boldsymbol{Y}_t'|\Sigma, D_{t-1}] = \boldsymbol{F}_t'\boldsymbol{a}_t$, $\mathrm{V}_l[\boldsymbol{Y}_t'|\Sigma, D_{t-1}] = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + V_t$, $\mathrm{V}_r[\boldsymbol{Y}_t'|\Sigma, D_{t-1}] = \Sigma$, where $\mathrm{V}_l[\cdot]$, $\mathrm{V}_r[\cdot]$ denote the left and right variances respectively. Use Theorem 4.4 to derive the marginal distribution. From Appendices A.2, B.2 we have that

$$\begin{aligned} \mathrm{C}[\mathrm{vec}(\Theta_t), \mathrm{vec}(\boldsymbol{Y}_t')|\Sigma, D_{t-1}] &= \Sigma \otimes (\boldsymbol{R}_t\boldsymbol{F}_t) \\ &= \Sigma \otimes (\boldsymbol{A}_t Q_t), \end{aligned}$$

so that $\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_t Q_t^{-1}$.

Write down the conditional joint distribution of Θ_t and $\boldsymbol{Y}_t'$ given D_{t-1} and Σ as

$$\left(\begin{array}{c} \Theta_t \\ \boldsymbol{Y}_t' \end{array} \middle| \Sigma, D_{t-1} \right) \sim \mathrm{N}\left[\left(\begin{array}{c} \boldsymbol{a}_t \\ \boldsymbol{f}_t' \end{array} \right), \left(\begin{array}{cc} \boldsymbol{R}_t & \boldsymbol{A}_t Q_t \\ \boldsymbol{A}_t' Q_t & Q_t \end{array} \right), \Sigma \right].$$

Applying the conditional results of the normal distribution (equation (B.2)) we have

$$(\Theta_t|\Sigma, D_t) \sim \mathrm{N}[\boldsymbol{m}_t, \boldsymbol{C}_t, \Sigma], \tag{4.17}$$

with quantities $\boldsymbol{m}_t$, $\boldsymbol{C}_t$ as stated in the Theorem. Further, write down the joint distribution of $\boldsymbol{Y}_t'$ and Σ given D_{t-1}

$$\left(\begin{array}{c} \boldsymbol{Y}_t' \\ \Sigma \end{array} \middle| D_{t-1} \right) \sim \mathrm{NGW}^{-1}[\boldsymbol{f}_t', Q_t, \boldsymbol{S}_{t-1}, \boldsymbol{N}_{t-1}, m_{t-1}].$$

So applying Theorem 4.2 we have

$$(\Sigma|D_t) \sim \mathrm{GW}^{-1}[\boldsymbol{S}_t, \boldsymbol{N}_t, m_t], \tag{4.18}$$

where

$$\boldsymbol{N}_t = \boldsymbol{N}_{t-1} + \boldsymbol{I} \qquad \text{and} \qquad \boldsymbol{N}_t^{1/2}\boldsymbol{S}_t\boldsymbol{N}_t^{1/2} = \boldsymbol{N}_{t-1}^{1/2}\boldsymbol{S}_{t-1}\boldsymbol{N}_{t-1}^{1/2} + \boldsymbol{e}_t Q_t^{-1}\boldsymbol{e}_t',$$

where $\boldsymbol{e}_t = \boldsymbol{Y}_t - \boldsymbol{f}_t$, and $m_t = r + \frac{\text{trace}(\boldsymbol{N}_t)}{2r}$. If we put together equations (4.17) and (4.18) we have the joint distribution of (d). The proof is complete by noting that from equation (4.16) (a) is true for $t = 1$. □

Identities

(1) $Q_t = (1 - \boldsymbol{F}'_t \boldsymbol{A}_t)^{-1} V_t$.

(2) $\boldsymbol{A}_t = \boldsymbol{C}_t \boldsymbol{F}_t / V_t$;

(3) $\boldsymbol{C}_t^{-1} = \boldsymbol{R}_t^{-1} + \boldsymbol{F}_t \boldsymbol{F}'_t / V_t$;

(4) $\boldsymbol{C}_t^{-1} \boldsymbol{m}_t = \boldsymbol{R}_t^{-1} \boldsymbol{a}_t + \boldsymbol{F}_t \boldsymbol{Y}'_t / V_t$.

Proof. The proof is exactly the same as in Section 2.3.2. □

The next theorem gives the k-step ahead forecast distributions.

Theorem 4.6. *For each time t and $k \geq 0$, the k-step ahead distributions for Θ_{t+k} and $\boldsymbol{Y}_{t+k}$ given D_t are given by*

(a) Joint state distribution: $(\Theta_{t+k}, \Sigma | D_t) \sim NGW^{-1}[\boldsymbol{a}_t(k), \boldsymbol{R}_t(k), \boldsymbol{S}_t, \boldsymbol{N}_t, m_t]$,

(b) Forecast distribution: $(\boldsymbol{Y}'_{t+k} | D_t) \sim GT[\boldsymbol{f}'_t, Q_t, \boldsymbol{S}_t, \boldsymbol{N}_t, p_t]$,

with moments defined recursively by

$$\boldsymbol{f}'_t = \boldsymbol{F}'_{t+k} \boldsymbol{a}_t(k), \qquad Q_t(k) = \boldsymbol{F}'_{t+k} \boldsymbol{R}_t(k) \boldsymbol{F}_{t+k} + V_{t+k},$$

where

$$\boldsymbol{a}_t(k) = \boldsymbol{G}_{t+k} \boldsymbol{a}_t(k-1), \qquad \boldsymbol{R}_t(k) = \boldsymbol{G}_{t+k} \boldsymbol{R}_t(k-1) \boldsymbol{G}'_{t+k} + \boldsymbol{W}_{t+k},$$

with starting values $\boldsymbol{a}_t(0) = \boldsymbol{m}_t$ and $\boldsymbol{R}_t(0) = \boldsymbol{C}_t$.

Proof. Define the $n \times n$ matrix $\boldsymbol{H}_{t+k}(x) = \boldsymbol{G}_{t+k}\boldsymbol{G}_{t+k-1}\dots\boldsymbol{G}_{t+k-x+1}$ for all t and integer $0 \leq x \leq k$, with $\boldsymbol{H}_{t+k}(0) = \boldsymbol{I}$. Note that $\boldsymbol{H}_{t+k}(x) = \boldsymbol{G}_{t+k}\boldsymbol{H}_{t+k-1}(x-1)$, $1 \leq x \leq k$. From repeated application of all the state evolution equations,

$$\Theta_{t+k} = \boldsymbol{H}_{t+k}(k)\Theta_t + \sum_{x=1}^{k} \boldsymbol{H}_{t+k}(k-x)\boldsymbol{\Omega}_{t+x},$$

from which it is deduced that

$$(\Theta_{t+k}|\Sigma, D_t) \sim \mathrm{N}[\boldsymbol{a}_t(k), \boldsymbol{R}_t(k), \Sigma],$$

with

$$\boldsymbol{a}_t(k) = \boldsymbol{H}_{t+k}(k)\boldsymbol{m}_t = \boldsymbol{G}_{t+k}\boldsymbol{a}_t(k-1)$$

and

$$\begin{aligned} \boldsymbol{R}_t(k) &= \boldsymbol{H}_{t+k}(k)\boldsymbol{C}_t\boldsymbol{H}'_{t+k}(k) + \sum_{x=1}^{k} \boldsymbol{H}_{t+k}(k-x)\boldsymbol{W}_{t+x}\boldsymbol{H}'_{t+k}(k-x) \\ &= \boldsymbol{G}_{t+k}\boldsymbol{R}_t(k-1)\boldsymbol{G}'_{t+k} + \boldsymbol{W}_{t+k}, \end{aligned}$$

with starting values $\boldsymbol{a}_t(0) = \boldsymbol{m}_t$ and $\boldsymbol{R}_t(0) = \boldsymbol{C}_t$. (a) follows, if the above conditional on Σ normal distribution is combined with

$$(\Sigma|D_t) \sim \mathrm{GW}^{-1}[\boldsymbol{S}_t, \boldsymbol{N}_t, m_t]. \tag{4.19}$$

Using the observation equation at time $t+k$ the forecast distribution conditional on Σ is deduced as $(\boldsymbol{Y}'_{t+k}|\Sigma, D_t) \sim \mathrm{N}[\boldsymbol{f}_t(k)', Q_t(k), \Sigma]$, where

$$\begin{aligned} \boldsymbol{f}_t(k)' &= \boldsymbol{F}'_{t+k}\boldsymbol{H}_{t+k}(k)\boldsymbol{m}_t = \boldsymbol{F}'_{t+k}\boldsymbol{a}_t(k), \quad \text{and} \\ Q_t(k) &= \boldsymbol{F}'_{t+k}\boldsymbol{R}_t(k)\boldsymbol{F}_{t+k} + V_{t+k}. \end{aligned}$$

Now (b) is derived by using this normal distribution and equation (4.19) to derive the marginal distribution $(\boldsymbol{Y}_{t+k}|D_t)$. □

Note that the notation of matrix normal distributions for $\boldsymbol{Y}_t$, $(t \geq 1)$, may look odd for a vector of observations. This is due to the generalization of the model to incorporate a matrix of observations, according to Quintana [34]. So if $\boldsymbol{Y}_t$ is an $r \times s$ matrix Theorems 4.5, 4.6 still hold as all the following results do with the modification that V_t, Q_t are replaced by the $s \times s$ variance matrices $\boldsymbol{V}_t$, $\boldsymbol{Q}_t$ respectively. Also the updating of the matrix of degrees of freedom changes to $\boldsymbol{N}_t = \boldsymbol{N}_{t-1} + s\boldsymbol{I}$. So the matrix version of the ECCM is

$$\boldsymbol{Y}_t' = \boldsymbol{F}_t'\Theta_t + \boldsymbol{\nu}_t', \qquad \boldsymbol{\nu}_t' \sim \mathrm{N}[\mathrm{O}, \boldsymbol{V}_t, \Sigma], \tag{4.20}$$

$$\Theta_t = \boldsymbol{G}_t\Theta_{t-1} + \boldsymbol{\Omega}_t, \qquad \boldsymbol{\Omega}_t \sim \mathrm{N}[\mathrm{O}, \boldsymbol{W}_t, \Sigma], \tag{4.20$'$}$$

$$(\Theta_0, \Sigma | D_0) \sim \mathrm{NGW}^{-1}[\boldsymbol{m}_0, \boldsymbol{C}_0, \boldsymbol{S}_0, \boldsymbol{N}_0, m_0], \tag{4.21}$$

where $\dim(\boldsymbol{Y}_t) = r \times s$, $\dim(\boldsymbol{F}_t) = n \times s$, $\dim(\Theta_t) = n \times r$, $\dim(\boldsymbol{\nu}_t) = r \times s$, $\boldsymbol{V}_t$ is a known $s \times s$ SPD matrix, and the remaining are defined as in model (4.15), (4.15$'$), and (4.16). The model is characterized by the common components $\boldsymbol{F}_t$, $\boldsymbol{G}_t$, $\boldsymbol{V}_t$, and $\boldsymbol{W}_t$. It can be shown that this model is equivalent to a series of multivariate DLMs with common components and so is particularly limited. Furthermore, this model is embedded in the GMDLM (General Multivariate DLM). To see this, set

$$\boldsymbol{Y}_t^* = \mathrm{vec}(\boldsymbol{Y}_t'), \qquad \boldsymbol{F}_t^* = \boldsymbol{I}_r \otimes \boldsymbol{F}_t, \qquad \boldsymbol{\theta}_t^* = \mathrm{vec}(\Theta_t), \tag{4.22}$$

$$\boldsymbol{\nu}_t^* = \mathrm{vec}(\boldsymbol{\nu}_t'), \qquad \boldsymbol{G}_t^* = \boldsymbol{I}_r \otimes \boldsymbol{G}_t, \qquad \boldsymbol{\omega}_t^* = \mathrm{vec}(\boldsymbol{\Omega}_t). \tag{4.22$'$}$$

Now model (4.20), (4.20$'$), (4.21) can take the form

$$\boldsymbol{Y}_t^* = \boldsymbol{F}_t^{*\prime}\boldsymbol{\theta}_t^* + \boldsymbol{\nu}_t^*, \qquad \boldsymbol{\nu}_t^* \sim \mathrm{N}[0, \Sigma \otimes \boldsymbol{V}_t],$$

$$\boldsymbol{\theta}_t^* = \boldsymbol{G}_t^*\boldsymbol{\theta}_{t-1}^* + \boldsymbol{\omega}_t^*, \qquad \boldsymbol{\omega}_t^* \sim \mathrm{N}[0, \Sigma \otimes \boldsymbol{W}_t],$$

and $\Sigma \sim \mathrm{GW}^{-1}[\boldsymbol{S}_0, \boldsymbol{N}_0, m_0]$. This makes clear that the above Matrix DLMs are subclasses of the GMDLM.

Returning to the ECCM, the next result gives the posterior distributions of $\boldsymbol{\nu}_t$ and $\boldsymbol{\Omega}_t$, given D_t.

Theorem 4.7. *In the ECCM define*

$$U_t = 1 - V_t Q_t^{-1}$$

$$\boldsymbol{L}_t = (1 - U_t)\boldsymbol{F}_t' \boldsymbol{W}_t$$

$$\boldsymbol{H}_t = \boldsymbol{W}_t - \boldsymbol{W}_t \boldsymbol{F}_t Q_t^{-1} \boldsymbol{F}_t' \boldsymbol{W}_t.$$

Then, given D_t, the posterior joint distribution of $(\boldsymbol{\nu}_t, \boldsymbol{\Omega}_t')'$ and Σ is

$$\left(\begin{pmatrix} \boldsymbol{\nu}_t' \\ \boldsymbol{\Omega}_t \end{pmatrix}, \Sigma \middle| D_t \right) \sim NGW^{-1}\left[\begin{pmatrix} 1 - U_t \\ \boldsymbol{W}_t \boldsymbol{F}_t Q_t^{-1} \end{pmatrix} \boldsymbol{e}_t', \begin{pmatrix} U_t V_t & -\boldsymbol{L}_t \\ -\boldsymbol{L}_t' & \boldsymbol{H}_t \end{pmatrix}, \boldsymbol{S}_t, \boldsymbol{N}_t, m_t \right].$$

Proof. First write down the joint distribution

$$\left(\begin{matrix} \boldsymbol{\nu}_t' \\ \boldsymbol{\Omega}_t \\ \boldsymbol{Y}_t' \end{matrix} \middle| \Sigma, D_{t-1} \right) \sim \mathrm{N}\left[\begin{pmatrix} \mathrm{0}' \\ \mathrm{O} \\ \boldsymbol{f}_t' \end{pmatrix}, \begin{pmatrix} V_t & \mathrm{O} & V_t \\ \mathrm{O}' & \boldsymbol{W}_t & \boldsymbol{W}_t \boldsymbol{F}_t \\ V_t & \boldsymbol{F}_t' \boldsymbol{W}_t & Q_t \end{pmatrix}, \Sigma \right].$$

Using standard normal conditional results (Appendix B.2)

$$\left(\begin{matrix} \boldsymbol{\nu}_t' \\ \boldsymbol{\Omega}_t \end{matrix} \middle| \Sigma, D_t \right) \sim \mathrm{N}\left[\begin{pmatrix} 1 - U_t \\ \boldsymbol{W}_t \boldsymbol{F}_t Q_t^{-1} \end{pmatrix} \boldsymbol{e}_t', \begin{pmatrix} U_t V_t & -\boldsymbol{L}_t \\ -\boldsymbol{L}_t' & \boldsymbol{H}_t \end{pmatrix}, \Sigma \right],$$

with U_t, $\boldsymbol{L}_t$, $\boldsymbol{H}_t$ as defined in the theorem. The distribution $(\Sigma | D_t) \sim \mathrm{GW}^{-1}[\boldsymbol{S}_t, \boldsymbol{N}_t, m_t]$ completes the proof. □

Consider the model (4.15), (4.15′), (4.16) with the partition

$$Y'_t = (Y'_{1t}, Y'_{2t}), \qquad \Theta_t = (\Theta_{1t}, \Theta_{2t}), \qquad \nu'_t = (\nu'_{1t}, \nu'_{2t}),$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix}, \qquad \Omega_t = (\Omega_{1t}, \Omega_{2t}),$$

where $\dim(Y_{1t}) = \dim(\nu_{1t}) = q \times 1$, $\dim(Y_{2t}) = \dim(\nu_{2t}) = (r-q) \times 1$, $\dim(\Theta_{1t}) = \dim(\Omega_{1t}) = n \times q$, $\dim(\Theta_{2t}) = \dim(\Omega_{2t}) = n \times (r-q)$, $\dim(\Sigma_{11}) = q \times q$, $\dim(\Sigma_{12}) = q \times (r-q)$, and $\dim(\Sigma_{22}) = (r-q) \times (r-q)$, for some $1 \leq q \leq r-1$.

Also let $D_{1t} = \{D_{1,t-1}, Y_{1t}\}$ and $D_{2t} = \{D_{2,t-1}, Y_{2t}\}$ represent the respective sets of information, so that $D_t = D_{1t} \bigcup D_{2t}$.

Referring to Theorem 4.5, write

$$m_t = (m_{1t}, m_{2t}), \qquad S_t = \begin{pmatrix} S_{11,t} & S_{12,t} \\ S'_{12,t} & S_{22,t} \end{pmatrix}, \qquad N_t = \begin{pmatrix} N_{11,t} & \mathrm{O} \\ \mathrm{O}' & N_{22} \end{pmatrix},$$

where $\dim(m_{1t}) = n \times q$, $\dim(m_{2t}) = n \times (r-q)$, $\dim(S_{11,t}) = \dim(N_{11,t}) = q \times q$, $\dim(S_{12,t}) = q \times (r-q)$, and $\dim(S_{22,t}) = \dim(N_{22,t}) = (r-q) \times (r-q)$.

Thus, the model of (4.15), (4.15′), (4.16) can be written as

$$\begin{aligned}
Y'_{1t} &= F'_t \Theta_{1t} + \nu'_{1t}, & \nu'_{1t} &\sim \mathrm{N}[0, V_t, \Sigma_{11}], \\
\Theta_{1t} &= G_t \Theta_{1,t-1} + \Omega_{1t}, & \Omega_{1t} &\sim \mathrm{N}[\mathrm{O}, W_t, \Sigma_{11}], \\
Y'_{2t} &= F'_t \Theta_{2t} + \nu'_{2t}, & \nu'_{2t} &\sim \mathrm{N}[0, V_t, \Sigma_{22}], \\
\Theta_{2t} &= G_t \Theta_{2,t-1} + \Omega_{2t}, & \Omega_{2t} &\sim \mathrm{N}[\mathrm{O}, W_t, \Sigma_{22}],
\end{aligned}$$

and

$$(\Theta_{1,0}, \Sigma | D_{1,0}) \sim \mathrm{NGW}^{-1}[m_{1,0}, C_0, S_{11,0}, N_{11,0}, m_{1,0}],$$

$$(\Theta_{2,0}, \Sigma | D_{2,0}) \sim \mathrm{NGW}^{-1}[m_{2,0}, C_0, S_{22,0}, N_{22,0}, m_{2,0}],$$

with $\mathrm{C}[\boldsymbol{\nu}_{1t}, \boldsymbol{\nu}_{2t} | \boldsymbol{\Sigma}, D_{t-1}] = \boldsymbol{\Sigma}_{12}$, for some known quantities $\boldsymbol{m}_{i,0}, \boldsymbol{S}_{ii,0}, \boldsymbol{N}_{ii,0}, \boldsymbol{C}_0$, $(i = 1, 2)$.

If at time t we know the values of $\boldsymbol{Y}_{1,t+1}, \ldots, \boldsymbol{Y}_{1,t+k}$, from Theorem 4.6 and equation (4.14) we have

$$(\boldsymbol{Y}'_{2,t+k} | D_{1,t+k}, D_t) \sim \mathrm{GT}[\boldsymbol{f}'_{2|1,t}(k), \boldsymbol{Q}_{2t}(k), \boldsymbol{S}_{2|1,t}, \boldsymbol{N}_{22,t}, p_1],$$

where

$$\boldsymbol{f}'_{2|1,t}(k) = \boldsymbol{f}'_{2t}(k) + (\boldsymbol{Y}'_{1,t+k} - \boldsymbol{f}'_{1t}(k))\boldsymbol{N}_{11,t}^{-1/2}\boldsymbol{S}_{11,t}^{-1}\boldsymbol{S}_{12,t}\boldsymbol{N}_{22,t}^{1/2},$$

$$\boldsymbol{Q}_{2t}(k) = \boldsymbol{Q}_t(k) + (\boldsymbol{Y}'_{1,t+k} - \boldsymbol{f}'_{1,t+k})\boldsymbol{N}_{11,t}^{-1/2}\boldsymbol{S}_{11,t}^{-1}\boldsymbol{N}_{11,t}^{-1/2}(\boldsymbol{Y}_{1,t+k} - \boldsymbol{f}_{1,t+k}),$$

$$\boldsymbol{S}_{2|1,t} = \boldsymbol{S}_{22,t} - \boldsymbol{S}'_{12,t}\boldsymbol{S}_{11,t}^{-1}\boldsymbol{S}_{12,t}, \qquad p_1 = q + \frac{\mathrm{trace}(\boldsymbol{N}_{22,t})}{r},$$

with

$$\boldsymbol{f}'_{1t} = \boldsymbol{F}'_{t+k}\boldsymbol{m}_{1t}, \qquad \boldsymbol{f}'_{2t}(k) = \boldsymbol{F}'_{t+k}\boldsymbol{m}_{2t},$$

and $\boldsymbol{Q}_t(k)$ as in Theorem 4.6.

This may find an application to multivariate time series, at least one marginal series of which is explicitly known for some time ahead. As an example, let Y_{1t} be the price index and Y_{2t} be the increase of the average standard salary. Write $\boldsymbol{Y}_t = (Y_{1t}, Y_{2t})'$. According to national and international contracts, the price index is known, at least for a couple of years ahead. Then, the conditional distribution of $Y_{2,t+k}$ given the known values of $Y_{1,t+k}$ and D_t, is obtained from the above analysis.

4.5 Retrospective Analysis

4.5.1 The ECCM Retrospective Parametric Distribution

Consider the ECCM (equations (4.15), (4.15′), (4.16)) and write the parameter matrix at time t as $\Theta_t = (\boldsymbol{\theta}_{t1}, \dots, \boldsymbol{\theta}_{tr})$, where each of the $n \times 1$ parameter vectors $\boldsymbol{\theta}_{ti}$, $(i = 1, \dots, r)$, corresponds to the r univariate DLMs. Since $\boldsymbol{A}_t$ is common throughout the r series, we define the $n \times n$ vector $\boldsymbol{A}_{t-k,t-j}$ as the common regression matrix of $\boldsymbol{\theta}_{t-k,i}$ on $\boldsymbol{\theta}_{t-j,i}$. Further, $\boldsymbol{B}_{t-k}$ is defined as $\boldsymbol{B}_{t-k} = \boldsymbol{C}_{t-k}\boldsymbol{G}'_{t-k+1}\boldsymbol{R}^{-1}_{t-k+1}$ and Θ_t, $\boldsymbol{Y}_{t+1}$ are conditionally independent given Θ_{t+1} and Σ, written as $\Theta_t \perp\!\!\!\perp \boldsymbol{Y}_{t+1}|\Theta_{t+1}, \Sigma$. Write $\Theta_t = \{\theta_{t,ij}\}$, $(i = 1, \dots, n; j = 1, \dots, r)$.

Theorem 4.8. *Given D_t, the joint distribution of the historical parameters $\Theta_1, \dots, \Theta_t$ and the variance Σ are defined by their marginal distributions and covariances*

$$(\Theta_{t-k}, \Sigma|D_t) \sim NGW^{-1}[\boldsymbol{a}_t(-k), \boldsymbol{R}_t(-k), \boldsymbol{S}_t, \boldsymbol{N}_t, m_t],$$

$$C[\theta_{t-k-j,pq}, \theta_{t-k,su}|\Sigma, D_t] = c^{(j)}_{t,ps}(-k)\sigma_{qu},$$

for $k \geq 0, 0 \leq j \leq t-k-1$ and $1 \leq p, s \leq n, 1 \leq q, u \leq r$, where the moments $\boldsymbol{a}_t(-k)$, $\boldsymbol{R}_t(-k)$ can be either calculated by

(i)

$$\boldsymbol{a}_t(-k) = \boldsymbol{a}_{t-1}(-k+1) + \boldsymbol{A}_{t-k,t}\boldsymbol{A}_t\boldsymbol{e}'_t,$$

$$\boldsymbol{R}_t(-k) = \boldsymbol{R}_{t-1}(-k+1) - \boldsymbol{A}_{t-k,t}\boldsymbol{A}_tQ_t\boldsymbol{A}'_t\boldsymbol{A}'_{t-k,t},$$

$$\boldsymbol{A}_{t-k,t} = \boldsymbol{A}_{t-k,t-1}\boldsymbol{B}_{t-1}.$$

(ii)

$$a_t(-k) = m_{t-k} + B_{t-k}[a_t(-k+1) - a_{t-k+1}],$$

$$R_t(-k) = C_{t-k} + B_{t-k}[R_t(-k+1) - R_{t-k+1}]B'_{t-k},$$

where $A_{t-k-j,t-k}R_t(-k) = \{c^{(j)}_{t,ps}(-k)\}$, $(p, s = 1, \ldots, n)$, $\Sigma = \{\sigma_{qu}\}$, $(q, u = 1, \ldots, r)$, $B_i = C_i G'_{i+1} R^{-1}_{i+1}$, $A_{t-k,t} = \prod_{i=0}^{k+1} B_{t-k-i}$, *and with initial values* $a_t(0) = m_t$, $a_{t-1}(1) = a_t$, $R_t(0) = C_t$, $R_{t-1}(1) = R_t$.

Proof. First we prove that

$$(\Theta_{t-k}|\Sigma, D_t) \sim \mathrm{N}[a_t(-k), R_t(-k), \Sigma]. \tag{4.23}$$

Considering (ii) and using the transformation (4.22), (4.22′), the above distributions are the respective filtered marginal distributions of Section 2.3.1, see equation (2.6). Referring to this section and denoting with "*" the relevant defining components there, under the above transformation it is clear that

$$a^*_t(-k) = \mathrm{vec}(a_t(-k)) = \mathrm{vec}(m_{t-k} + B_{t-k}[a_t(-k+1) - a_{t-k+1}]).$$

Note that $C^*_t = R^*_t(0) = \Sigma \otimes C_t = \Sigma \otimes R_t(0)$. Now assume that $R^*_t(-k+1) = \Sigma \otimes R_t(-k+1)$ and see that

$$R^*_t(-k) = \Sigma \otimes R_t(-k) = \Sigma \otimes [C_{t-k} + B_{t-k}(R_t(-k+1) - R_{t-k+1})B'_{t-k}],$$

with B_t as defined in the theorem. This proves (ii). Hence by induction we have (4.23).

Now for (i), write

$$a_t(-k) - a_{t-1}(-k+1) = B_{t-k}[a_t(-k+1) - a_{t-1}(-k+2)],$$

$$R_t(-k) - R_{t-1}(-k+1) = B_{t-k}[R_t(-k+1) - R_{t-1}(-k+2)]B'_{t-k},$$

which if applied recurrently provide the updatings of (ii).

These prove equation (4.23) with moments as in (i), (ii). The required marginal distribution of the theorem is straightforward by noting that $(\Sigma|D_t) \sim \mathrm{GW}^{-1}[\boldsymbol{S}_t, \boldsymbol{N}_t, m_t]$ and writing the joint distribution of Θ_{t-k} and Σ, given D_t. The covariance structure of the nr parameters $\theta_{t,il}$, $(i = 1, \ldots, n; l = 1, \ldots, r)$ is an immediate result from $\mathrm{C}[\boldsymbol{\theta}^*_{t-k-j}, \boldsymbol{\theta}^*_{t-k}|\Sigma, D_t] = \Sigma \otimes (\boldsymbol{A}_{t-k-j,t-k}\boldsymbol{R}_t(-k))$. □

Equations (i) are used to update the whole history up to a point of time in the past, while equations (ii) focus on a specific point of time in the past.

4.5.2 Deleting Observations

Let $\boldsymbol{Y}$ be any $r \times 1$ observation vector, $\boldsymbol{Z}$ an $n \times r$ parameter matrix, $\boldsymbol{F}$ an $n \times 1$ design vector and Σ an $r \times r$ unknown variance matrix such that

$$(\boldsymbol{Y}'|\boldsymbol{Z}, \Sigma) \sim \mathrm{N}[\boldsymbol{F}'\boldsymbol{Z}, R_{y|z}, \Sigma],$$

$$(\boldsymbol{Z}|\boldsymbol{Y}, \Sigma) \sim \mathrm{N}[\boldsymbol{a}_{z|y}, \boldsymbol{R}_{z|y}, \Sigma],$$

$$(\Sigma|\boldsymbol{Y}) \sim \mathrm{GW}^{-1}[\boldsymbol{S}_y, \boldsymbol{N}_y, m_y],$$

with a joint distribution

$$\left(\begin{array}{c} \boldsymbol{Z} \\ \boldsymbol{Y}' \end{array} \Bigg| \Sigma \right) \sim \mathrm{N}\left[\left(\begin{array}{c} \boldsymbol{a}_z \\ \boldsymbol{F}'\boldsymbol{a}_z \end{array} \right), \left(\begin{array}{cc} \boldsymbol{R}_z & \boldsymbol{R}_z\boldsymbol{F} \\ \boldsymbol{F}'\boldsymbol{R}_z & R_y \end{array} \right), \Sigma \right],$$

where $R_y = \boldsymbol{F}'\boldsymbol{R}_z\boldsymbol{F} + R_{y|z}$, for some known quantities $\boldsymbol{a}_z$, $\boldsymbol{a}_{z|y}$, $R_{y|z}$, $\boldsymbol{R}_{z|y}$, $\boldsymbol{R}_z$, $\boldsymbol{S}_y$, $\boldsymbol{N}_y$, and $m_y = r + \frac{\mathrm{trace}(\boldsymbol{N}_y)}{2r}$.

Theorem 4.9. *Given the above distributional structure and notation, define*

$$\boldsymbol{d}' = \boldsymbol{Y}' - \boldsymbol{F}'\boldsymbol{a}_{z|y}, \qquad R_d = R_{y|z} - \boldsymbol{F}'\boldsymbol{R}_{z|y}\boldsymbol{F}, \qquad \text{and} \qquad \boldsymbol{A}_{zd} = \boldsymbol{R}_{z|y}\boldsymbol{F}R_d^{-1}.$$

The following results hold:

(1) The leverage $\boldsymbol{A}$, of $\boldsymbol{Y}$ on $\boldsymbol{Z}$, is calculable as $\boldsymbol{A} = \boldsymbol{R}_{z|y}\boldsymbol{F}R_{y|z}^{-1}$.

(2) Deleting $\boldsymbol{Y}$ we have

$$(\boldsymbol{Z}|\Sigma) \sim N[\boldsymbol{a}_z, \boldsymbol{R}_z, \Sigma], \quad \Sigma \sim GW^{-1}[\boldsymbol{S}, \boldsymbol{N}, m], \quad m = r + \frac{trace(\boldsymbol{N})}{2r},$$

where

$$\boldsymbol{a}_z = \boldsymbol{a}_{z|y} - \boldsymbol{A}_{zd}\boldsymbol{d}', \quad \boldsymbol{R}_z = \boldsymbol{R}_{z|y} + \boldsymbol{A}_{zd}R_d\boldsymbol{A}'_{zd}, \quad \boldsymbol{N} = \boldsymbol{N}_y - \boldsymbol{I},$$

and

$$\boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2} = \boldsymbol{N}_y^{1/2}\boldsymbol{S}_y\boldsymbol{N}_y^{1/2} - \boldsymbol{d}'R_d^{-1}\boldsymbol{d}.$$

(3) The jackknife forecast for $\boldsymbol{Y}$ is

$$\boldsymbol{Y}' \sim GT[\boldsymbol{F}'\boldsymbol{a}_z, R_y, \boldsymbol{S}, \boldsymbol{N}, p],$$

where $R_y = \boldsymbol{F}'\boldsymbol{R}_z\boldsymbol{F} + R_{y|z}$ and $p = \frac{trace(\boldsymbol{N})}{r}$.

Proof. First see that $p(\boldsymbol{Z}|\boldsymbol{Y}, \Sigma) = p(\boldsymbol{Z}|\boldsymbol{d}, \Sigma)$. Then from Appendix B.2 it is

$$\boldsymbol{a}_{z|y} = \boldsymbol{a}_z + \boldsymbol{A}_{zd}\boldsymbol{d}',$$
$$\boldsymbol{R}_{z|y} = \boldsymbol{R}_z - \boldsymbol{A}_{zd}R_d\boldsymbol{A}'_{zd},$$

which provide the required moments $\boldsymbol{a}_z$, $\boldsymbol{R}_z$ of (2). Also, similarly with the identities (2), (3) of page 21, it is shown

$$\boldsymbol{R}_{z|y}^{-1} = \boldsymbol{R}_z^{-1} + \boldsymbol{F}R_{y|z}^{-1}\boldsymbol{F}', \tag{4.24}$$
$$\boldsymbol{R}_{z|y}^{-1}\boldsymbol{a}_{z|y} = \boldsymbol{R}_z^{-1}\boldsymbol{a}_z + \boldsymbol{F}R_{y|z}^{-1}\boldsymbol{Y}',$$

from the later of which, the leverage is obtained as $\boldsymbol{A} = \boldsymbol{R}_{z|y}\boldsymbol{F}R_{y|z}^{-1}$. This establishes (1). Now equation (4.24) can be written as

$$\boldsymbol{R}_z = \boldsymbol{R}_{z|y} + \boldsymbol{R}_{z|y}\boldsymbol{F}R_{y|z}^{-1}\boldsymbol{F}'\boldsymbol{R}_z,$$

from which we get $\boldsymbol{A} = \boldsymbol{R}_z\boldsymbol{F}R_y^{-1}$. Also (4.24) yields

$$\boldsymbol{R}_{z|y}\boldsymbol{F} = \boldsymbol{R}_z\boldsymbol{F} - \boldsymbol{R}_z\boldsymbol{F}R_{y|z}^{-1}\boldsymbol{F}'\boldsymbol{R}_{z|y}\boldsymbol{F},$$

from which it is obtained $\boldsymbol{A}_{zd} = \boldsymbol{R}_z\boldsymbol{F}R_{y|z}^{-1}$.

To prove the remaining results of (2), first see that $\boldsymbol{A}_{zd}\boldsymbol{d}' = \boldsymbol{A}\boldsymbol{e}'$, from the distribution of $(\boldsymbol{Z}|\boldsymbol{Y},\Sigma)$. Then,

$$\boldsymbol{e}' = \boldsymbol{Y}' - \boldsymbol{F}'\boldsymbol{a}_{z|y} + \boldsymbol{F}'(\boldsymbol{a}_{z|y} - \boldsymbol{a}_z) = \boldsymbol{d}' + \boldsymbol{F}'\boldsymbol{A}\boldsymbol{e}',$$

which implies

$$\boldsymbol{d}' = (1 - \boldsymbol{F}'\boldsymbol{A})\boldsymbol{e}' = (1 - \boldsymbol{F}'\boldsymbol{R}_z\boldsymbol{F}R_y^{-1})\boldsymbol{e}' = R_{y|z}R_y^{-1}\boldsymbol{e}'.$$

Also from $\boldsymbol{d}' = \boldsymbol{Y}' - \boldsymbol{F}'\boldsymbol{a}_z - \boldsymbol{F}'\boldsymbol{A}_{zd}\boldsymbol{d}'$

$$\begin{aligned} \boldsymbol{e}' &= (1 + \boldsymbol{F}'\boldsymbol{A}_{zd})\boldsymbol{d}' = (1 + \boldsymbol{F}'\boldsymbol{R}_{z|y}\boldsymbol{F}R_d^{-1})\boldsymbol{d}' \\ &= (R_d + \boldsymbol{F}'\boldsymbol{R}_{z|y}\boldsymbol{F})R_d^{-1}\boldsymbol{d}' = R_{y|z}R_d^{-1}\boldsymbol{d}'. \end{aligned}$$

Now using the identities

$$\boldsymbol{N}_y = \boldsymbol{N} + \boldsymbol{I}, \qquad \boldsymbol{d}' = R_{y|z}R_y^{-1}\boldsymbol{e}', \qquad \text{and} \qquad \boldsymbol{e}' = R_{y|z}R_d^{-1}\boldsymbol{d}',$$

we see that

$$\begin{aligned} \boldsymbol{N}_y^{1/2}\boldsymbol{S}_y\boldsymbol{N}_y^{1/2} &= \boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2} + \boldsymbol{e}R_y^{-1}\boldsymbol{e}' \\ &= \boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2} + \boldsymbol{d}R_{y|z}^{-1}R_{y|z}R_d^{-1}\boldsymbol{d}' \\ &= \boldsymbol{N}^{1/2}\boldsymbol{S}\boldsymbol{N}^{1/2} + \boldsymbol{d}R_d^{-1}\boldsymbol{d}'. \end{aligned}$$

Then $N^{1/2}SN^{1/2} = N_y^{1/2}S_yN_y^{1/2} - dR_d^{-1}d'$ and $N = N_y - I$, completing (2). The jackknife forecast of (3) now follows immediately from Theorem 4.4. □

An advantage of using the generalized inverse Wishart distribution in this theorem is a minimal assumption $n_{y,i} > 1$ ($N_y = \text{diag}\{n_{y1}, \ldots, n_{yr}\}$ for every $i = 1, \ldots, r$), while in any other multivariate formulation using either inverse Wishart or inverse gamma distributions it must be $n_y > r$ so that $n = n_y - r$ is always positive, see [20] and [51, page 118]. This means that the incorporation of a matrix of degrees of freedom in the models allows for a significantly faster and more effective deletion of observations, especially with large r.

4.5.3 Deleting Observations in the ECCM

Write $D_t(-k) = \{D_t - Y_{t-k}\}$ to be the current information except for the observed value of Y_{t-k}. Given D_t, define the following retrospective quantities relating to time $t-k$ as

$$
\begin{aligned}
e_t'(-k) &= Y_{t-k}' - F_{t-k}'a_t(-k), \\
Q_t(-k) &= V_{t-k} - F_{t-k}'R_t(-k)F_{t-k}, \\
A_t(-k) &= R_t(-k)F_{t-k}[Q_t(-k)]^{-1}.
\end{aligned}
$$

Theorem 4.10. *Deleting the observation Y_{t-k}, the following distributional results hold*

(*i*) $\quad (\Theta_{t-k}, \Sigma | D_t(-k)) \sim NGW^{-1}[a_{t,k}, R_{t,k}, S_{t,k}, N, m]$,

(*ii*) $\quad (Y_{t-k}' | D_t(-k)) \sim GT[f_{t,k}', Q_{t,k}, S_{t,k}, N, p]$,

where

$$a_{t,k} = a_t(-k) - A_t(-k)e_t'(-k), \qquad f_{t,k}' = F_{t-k}'a_{t,k},$$

$$R_{t,k} = R_t(-k) + A_t(-k)Q_t(-k)A_t'(-k), \qquad Q_{t,k} = F_{t-k}'R_{t,k}F_{t-k} + V_{t-k},$$

$$N^{1/2}S_{t,k}N^{1/2} = N_t^{1/2}S_tN_t^{1/2} - e_t(-k)[Q_t(-k)]^{-1}e_t'(-k),$$

and

$$N = N_t - I, \qquad m = r + \frac{trace(N)}{2r}, \qquad p = \frac{trace(N)}{r}.$$

Proof. This is proven by applying Theorem 4.9, with settings $Z = \Theta_{t-k}$, $Y = Y_{t-k}$, $R_d = Q_t(-k)$, $R_{y|z} = V_{t-k}$, $d = e_t(-k)$, and $A_{zd} = A_t(-k)$, with the joint distribution of (Z, Y) being (implicitly) conditional on $D_t(-k)$. □

The distribution of (ii) is the jackknife forecast distribution of Y_{t-k}, and the residual $e_{t,k} = Y_{t-k} - f_{t,k}$ is the jackknife residual of this observation.

A graph of all the standardized jackknife residuals is particularly useful in detecting influential observations, thus contributing to model assessment, see [23].

The influence of the individual observation Y_{t-k} on the parametric mean Θ_{t-k} is measured with the leverage of Y_{t-k} as

$$R_{t,k}F_{t-k}(F_{t-k}'R_{t,k}F_{t-k} + V_{t-k})^{-1}.$$

Now in the framework of Theorem 4.9, assume that Y_1 is any $l \times 1$ subvector of Y, such that

$$(Y_1|Z_1, \Sigma_{11}) \sim \mathrm{N}[F'Z_1, R_{y1|z1}, \Sigma_{11}],$$

where the $n \times l$ matrix Z_1 comes from Z when we exclude the corresponding $r - l$ columns and the $l \times l$ SPD matrix Σ_{11} is such that

$$\Sigma_{11} \sim \mathrm{GW}^{-1}[S_{y11}, N_{y11}, m_{y1}],$$

where

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}, \quad S_y = \begin{pmatrix} S_{y11} & S_{y12} \\ S_{y12}' & S_{y22} \end{pmatrix}, \quad N_y = \begin{pmatrix} N_{y11} & O \\ O' & N_{y22} \end{pmatrix},$$

with $\dim(\Sigma_{11}) = \dim(S_{y11}) = \dim(N_{y1}) = l \times l$, $\dim(\Sigma_{12}) = \dim(S_{y12}) = l \times (r-l)$, $\dim(\Sigma_{22}) = \dim(S_{y22}) = \dim(N_{y22}) = (r-l) \times (r-l)$, and as usual $m_{y1} = l + \frac{\text{trace}(N_v)}{2r}$. The above distribution of Σ_{11} is obtained by Theorem 4.1. Notice that, whatever the order of the l single deleting observations of Y, we can always rearrange them so that we obtain the above structure. Hence, Theorem 4.9 is trivially extended to incorporate the subvector Y_1 instead of Y. Now for the ECCM write $D_{t,l}(-k) = \{D_t - Y_{t-k,l}\}$ to be the current information except for the observed $l \times 1$ subvector of Y_{t-k}, $Y_{t-k,l}$. Given D_t, define the following retrospective quantities

$$e_{t,l}'(-k) = Y_{t-k,l}' - F_{t-k}' a_{t,l}(-k),$$
$$Q_{t,l}(-k) = V_{t-k} - F_{t-k}' R_{t,l}(-k) F_{t-k},$$
$$A_{t,l}(-k) = R_{t,l}(-k) F_{t-k} [Q_{t,l}(-k)]^{-1}.$$

Theorem 4.11. *Deleting the observation $Y_{t-k,l}$, the following distributional results hold*

(*i*) $\quad (\Theta_{t-k,l}, \Sigma_{11} | D_{t,l}(-k)) \sim NGW^{-1}[a_{t,k,l}, R_{t,k,l}, S_{t,k,l}, N_{11}, m_1],$

(*ii*) $\quad (Y_{t-k,l}' | D_{t,l}(-k)) \sim GT[f_{t,k,l}', Q_{t,k,l}, S_{t,k,l}, N_{11}, p_1],$

where

$$a_{t,k,l} = a_{t,l}(-k) - A_{t,l}(-k) e_{t,l}'(-k), \qquad f_{t,k,l}' = F_{t-k}' a_{t,k,l},$$
$$R_{t,k,l} = R_{t,l}(-k) + A_{t,l}(-k) Q_{t,l}(-k) A_{t,l}'(-k), \quad Q_{t,k,l} = F_{t-k}' R_{t,k,l} F_{t-k} + V_{t-k},$$
$$N_{11}^{1/2} S_{t,k,l} N_{11}^{1/2} = N_{11,t}^{1/2} S_t N_{11,t}^{1/2} - e_{t,l}(-k) [Q_{t,l}(-k)]^{-1} e_{t,l}'(-k),$$

and

$$\boldsymbol{N}_{11} = \boldsymbol{N}_{11,t} - \boldsymbol{I}, \qquad m_1 = l + \frac{trace(\boldsymbol{N})}{2r}, \qquad p_1 = l - r + \frac{trace(\boldsymbol{N})}{r},$$

where $\boldsymbol{N} = diag\{\boldsymbol{N}_{11}, \boldsymbol{N}_{22}\}$, $\boldsymbol{N}_t = diag\{\boldsymbol{N}_{11,t}, \boldsymbol{N}_{22,t}\}$, *and* $\boldsymbol{S}_{t,k} = \{\boldsymbol{S}_{t,k,ij}\}$, $i, j = 1, 2$.

Proof. Again using Theorem 4.9, with settings $\boldsymbol{Z} = \Theta_{t-k,l}$, $\boldsymbol{Y} = \boldsymbol{Y}_{t-k,l}$, $R_d = Q_{t,l}(-k)$, $R_{y|z} = V_{t-k}$, $\boldsymbol{d} = \boldsymbol{e}_{t,l}(-k)$, $\boldsymbol{A}_{zd} = \boldsymbol{A}_{t,l}(-k)$, we derive the conditional distribution of $(\Theta_{t-k,l}|\Sigma_{11}, D_{t,l}(-k))$ which with Theorem 4.1 provides the required equations. □

The leverage of the observation $\boldsymbol{Y}_{t-k,l}$ is

$$\boldsymbol{R}_{t,k,l}\boldsymbol{F}_{t-k}(\boldsymbol{F}'_{t-k}\boldsymbol{R}_{t,k,l}\boldsymbol{F}_{t-k} + V_{t-k})^{-1}.$$

4.6 Reference Analysis

Practitioners very often have difficulty specifying the starting values $\boldsymbol{m}_0$, $\boldsymbol{C}_0$. More importantly there is a problem of identifying proper prior distributions. Most practitioners employ a normal generalized inverse Wishart prior, say $(\Theta, \Sigma) \sim \text{NGW}^{-1}[\boldsymbol{m}_0, \boldsymbol{C}_0, \boldsymbol{S}_0, m_0]$. Usually they set $\boldsymbol{S}_0 = \boldsymbol{N}_0 = \boldsymbol{I}$, $m_0 = r + 1/2$, and, according to history, they specify $\boldsymbol{m}_0$ and $\boldsymbol{C}_0$. The posterior distributions, hence all the model future depends hugely upon these proper priors and their corresponding posterior distributions. This is not a new problem. Bayesian Statistics has been criticized due to the possible miss-specification of these initial distributions. Pole and West ([30]) have solved the case of the univariate DLM for both the known and unknown variances.

Assuming an initial Jeffreys prior for the states they derive sequential updating for all t (prior and posterior of the states as long as priors remain improper) including the linking points when the priors become proper. Barbosa ([3, chapter 5]) discussed the problem of the CCM and the multivariate case with known variances. However, the problem of missing observations and intervention for the CCM still remains unsolved. According to [3], the distributions remain improper at least until the first $n + (r+1)/2$ observations. So it is very likely that partial missing observations (definition to follow in Chapter 6), will cause problems. We solve the complete problem of missing observations and variance intervention by developing the general theory, which includes the existing univariate and CCM reference analysis.

Theorem 4.12. *For the ECCM let the initial prior distribution be represented by the reference form*

$$p(\Theta_1, \Sigma | D_0) \propto |\Sigma|^{-(r+1)/2}.$$

Then, assuming that $\boldsymbol{W}_t$ has full rank,

(1) The joint prior and posterior distributions of the parameter Θ_t and the variance Σ are given by

$$\begin{aligned} p(\Theta_t, \Sigma | D_{t-1}) &\propto |\Sigma|^{-\frac{1}{2}\left(\frac{trace(\boldsymbol{\Gamma}_{t-1})}{r}+r+1\right)} exp\{-\frac{1}{2} trace\{[\Theta_t' \boldsymbol{H}_t \Theta_t \\ &\quad -2\Theta_t' \boldsymbol{h}_t + \Lambda_t]\Sigma^{-1}\}\}, \\ p(\Theta_t, \Sigma | D_t) &\propto |\Sigma|^{-\frac{1}{2}\left(\frac{trace(\boldsymbol{\Gamma}_t)}{r}+r+1\right)} exp\{-\frac{1}{2} trace\{[\Theta_t' \boldsymbol{K}_t \Theta_t \\ &\quad -2\Theta_t' \boldsymbol{k}_t + \Delta_t]\Sigma^{-1}\}\}, \end{aligned}$$

where

$$\boldsymbol{H}_t = \boldsymbol{W}_t^{-1} - \boldsymbol{W}_t^{-1}\boldsymbol{G}_t\boldsymbol{P}_t^{-1}\boldsymbol{G}_t'\boldsymbol{W}_t^{-1},$$

$$\boldsymbol{P}_t = \boldsymbol{G}_t'\boldsymbol{W}_t^{-1}\boldsymbol{G}_t + \boldsymbol{K}_{t-1},$$

$$\boldsymbol{h}_t = \boldsymbol{W}_t^{-1}\boldsymbol{G}_t\boldsymbol{P}_t^{-1}\boldsymbol{k}_{t-1},$$

$$\boldsymbol{K}_t = \boldsymbol{H}_t + \boldsymbol{F}_t V_t^{-1}\boldsymbol{F}_t',$$

$$\boldsymbol{k}_t = \boldsymbol{h}_t + \boldsymbol{F}_t V_t^{-1}\boldsymbol{Y}_t',$$

$$\boldsymbol{\Lambda}_t = \boldsymbol{\Delta}_{t-1} - \boldsymbol{k}_{t-1}'\boldsymbol{P}_t^{-1}\boldsymbol{k}_{t-1},$$

$$\boldsymbol{\Delta}_t = \boldsymbol{\Lambda}_t + \boldsymbol{Y}_t V_t^{-1}\boldsymbol{Y}_t',$$

$$\boldsymbol{\Gamma}_t = \boldsymbol{\Gamma}_{t-1} + \boldsymbol{I},$$

with starting values $\boldsymbol{H}_1 = \mathbf{O}$, $\boldsymbol{h}_1 = \mathbf{O}$, $\boldsymbol{\Lambda}_1 = \mathbf{O}$, *and* $\boldsymbol{\Gamma}_0 = \mathbf{O}$.

(2) For $t \geq t_{n,r}$, *where* $t_{n,r} = n + (r+1)/2$, *if no missing observations exist until* $t = t_{n,r}$, *the posterior distribution* $(\Theta_t, \Sigma | D_t)$ *is a matrix normal generalized inverse Wishart distribution*

$$(\Theta_t, \Sigma | D_t) \sim NGW^{-1}[\boldsymbol{m}_t, \boldsymbol{C}_t, \boldsymbol{S}_t, \boldsymbol{N}_t, m_t],$$

where

$$\boldsymbol{m}_t = \boldsymbol{K}_t^{-1}\boldsymbol{k}_t, \qquad \boldsymbol{C}_t = \boldsymbol{K}_t^{-1},$$

$$\boldsymbol{N}_t^{1/2}\boldsymbol{S}_t\boldsymbol{N}_t^{1/2} = \boldsymbol{\Delta}_t - \boldsymbol{k}_t'\boldsymbol{m}_t, \qquad \boldsymbol{N}_t = \boldsymbol{\Gamma}_t - (r+n-1)\boldsymbol{I},$$

and as usual $m_t = r + \dfrac{trace(\boldsymbol{N}_t)}{2r}$.

Proof. The proof is by induction. Assume that at time t the prior distribution of $(\Theta_t, \Sigma | D_{t-1})$ is true,

$$p(\Theta_t, \Sigma | D_{t-1}) \;\propto\; |\Sigma|^{-\frac{1}{2}\left(\frac{\mathrm{trace}(\boldsymbol{\Gamma}_{t-1})}{r}+r+1\right)}\exp\{-\frac{1}{2}\mathrm{trace}\{[\Theta_t'\boldsymbol{H}_t\Theta_t - 2\Theta_t'\boldsymbol{h}_t + \boldsymbol{\Lambda}_t]\Sigma^{-1}\}\}.$$

The likelihood from the observation $\boldsymbol{Y}_t$ is

$$
\begin{aligned}
p(\boldsymbol{Y}_t'|\Theta_t,\Sigma,D_{t-1}) &\propto |\Sigma|^{-1/2}\exp\{-\frac{1}{2}\mathrm{trace}\{[\Theta_t'\boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'\Theta_t - 2\Theta_t'\boldsymbol{F}_tV_t^{-1}\boldsymbol{Y}_t' \\
&\quad +\boldsymbol{Y}_tV_t^{-1}\boldsymbol{Y}_t']\Sigma^{-1}\}\}.
\end{aligned}
$$

By Bayes' theorem, the joint posterior distribution is

$$
\begin{aligned}
p(\Theta_t,\Sigma|D_t) &\propto p(\Theta_t,\Sigma|D_{t-1})p(\boldsymbol{Y}_t'|\Theta_t,\Sigma,D_{t-1}) \\
&= |\Sigma|^{-\frac{1}{2}\left(\frac{\mathrm{trace}(\boldsymbol{\Gamma}_t)}{r}+r+1\right)}\exp\{-\frac{1}{2}\mathrm{trace}\{[\Theta_t'\boldsymbol{K}_t\Theta_t - 2\Theta_t'\boldsymbol{k}_t \\
&\quad +\boldsymbol{\Delta}_t]\Sigma^{-1}\}\}, \qquad (4.25)
\end{aligned}
$$

with $\boldsymbol{K}_t$, $\boldsymbol{k}_t$, $\boldsymbol{\Delta}_t$, $\boldsymbol{\Gamma}_t$ as defined in the theorem. Now proceeding with induction, the joint prior distribution at time $t+1$ will be given by

$$
\begin{aligned}
p(\Theta_{t+1},\Sigma|D_t) &= \int p(\Theta_{t+1},\Sigma|\Theta_t,D_t)p(\Theta_t|D_t)\,d\Theta_t \\
&= \int p(\Theta_{t+1}|\Theta_t,\Sigma,D_t)p(\Theta_t,\Sigma|D_t)\,d\Theta_t.
\end{aligned}
$$

From the model evolution equation, the first term of the integrand part is a matrix normal distribution $\mathrm{N}[\boldsymbol{G}_{t+1}\Theta_t,\boldsymbol{W}_{t+1},\Sigma]$ and the second term is equation (4.25). Thus

$$
\begin{aligned}
p(\Theta_{t+1},\Sigma|D_t) &\propto \int |\Sigma|^{-\frac{1}{2}\left(\frac{\mathrm{trace}(\boldsymbol{\Gamma}_t)}{r}+r+n+1\right)}\exp\{-\frac{1}{2}\mathrm{trace}\{[\Theta_{t+1}'\boldsymbol{W}_{t+1}^{-1}\Theta_{t+1} \\
&\quad -2\Theta_t'\boldsymbol{G}_{t+1}'\boldsymbol{W}_{t+1}^{-1}\Theta_{t+1} + \Theta_t'\boldsymbol{G}_{t+1}'\boldsymbol{W}_{t+1}^{-1}\boldsymbol{G}_{t+1}\Theta_t + \Theta_t'\boldsymbol{K}_t\Theta_t \\
&\quad -2\Theta_t'\boldsymbol{k}_t + \boldsymbol{\Delta}_t]\Sigma^{-1}\}\}\,d\Theta_t \\
&= \int |\Sigma|^{-\frac{1}{2}\left(\frac{\mathrm{trace}(\boldsymbol{\Gamma}_t)}{r}+r+n+1\right)}\exp\{-\frac{1}{2}\mathrm{trace}\{[\Theta_t'\boldsymbol{P}_{t+1}\Theta_t \\
&\quad -2\Theta_t'\boldsymbol{\lambda}_t + \boldsymbol{\xi}_t]\Sigma^{-1}\}\}\,d\Theta_t \\
&= \int |\Sigma|^{-\frac{1}{2}\left(\frac{\mathrm{trace}(\boldsymbol{\Gamma}_t)}{r}+r+n+1\right)}\exp\{-\frac{1}{2}\mathrm{trace}\{[(\Theta_t-\boldsymbol{\alpha}_t)'\boldsymbol{P}_{t+1} \\
&\quad \times(\Theta_t-\boldsymbol{\alpha}_t) + \boldsymbol{\Xi}_t]\Sigma^{-1}\}\}\,d\Theta_t,
\end{aligned}
$$

where

$$\lambda_t = G'_{t+1}W_{t+1}^{-1}\Theta_{t+1} + k_t, \qquad \xi_t = \Theta'_{t+1}W_{t+1}^{-1}\Theta_{t+1} + \Delta_t,$$

$$\alpha_t = P_{t+1}^{-1}\lambda_t, \qquad \Xi_t = \xi_t - \alpha'_t P_{t+1}\alpha_t,$$

and P_t as defined in the theorem.

Integrating the matrix normal density we get

$$p(\Theta_{t+1}, \Sigma|D_t) \propto |\Sigma|^{-\frac{1}{2}\left(\frac{\text{trace}(\Gamma_t)}{r}+r+1\right)}\exp\{-\frac{1}{2}\text{trace}\{\Xi_t\Sigma^{-1}\}\}.$$

Now, substituting ξ_t and α_t in Ξ_t we have

$$\Xi_t = \Theta'_{t+1}H_{t+1}\Theta_{t+1} - 2\Theta'_{t+1}h_{t+1} + \Lambda_{t+1},$$

with H_{t+1}, h_{t+1}, Λ_{t+1} as required. To complete (1) it remains to validate the theorem for $t = 1$. Setting $H_1 = O$, $h_1 = O$, $\Lambda_1 = O$, and $\Gamma_0 = O$ provides a direct validation. Note that for $t = 1$ we get the joint prior distribution as the reference form

$$p(\Theta_1, \Sigma|D_0) \propto |\Sigma|^{-(r+1)/2}.$$

Considering now (2), if we assume that $t \geq t_{n,r}$ the posterior joint distributions are proper, hence the distribution of $(\Theta_t, \Sigma|D_t)$ can be written as

$$\begin{aligned} p(\Theta_t, \Sigma|D_t) \;\; \propto \;\; & |\Sigma|^{-\frac{n}{2}}\exp\{-\frac{1}{2}\text{trace}\{(\Theta_t - m_t)'C_t^{-1}(\Theta_t - m_t)\Sigma^{-1}\}\} \\ & \times|\Sigma|^{-\frac{1}{2}\left(\frac{\text{trace}(\Gamma_t)}{r}+r-n+1\right)}\exp\{-\frac{1}{2}\text{trace}\{[\Delta_t - k'_t m_t]\Sigma^{-1}\}\}, \end{aligned}$$

where m_t and C_t are identified from the matrix normal component as

$$m_t = K_t^{-1}k_t \qquad \text{and} \qquad C_t = K_t^{-1}.$$

Also, from the generalized inverse Wishart component, we have

$$N_t^{1/2}S_tN_t^{1/2} = \Delta_t - k'_t m_t \qquad \text{and} \qquad N_t = \Gamma_t - (r+n-1)I,$$

as required. □

Note that the reference form $p(\Theta_1, \Sigma | D_0) \propto |\Sigma|^{-(r+1)/2}$ can be derived from the Jeffreys prior, see [5, chapter 1], or [3, chapter 5].

Missing observations are fully considered in Chapter 6. The above theorem requires the existence of $\boldsymbol{W}_t^{-1}$ for every t. Since for any $0 < t < t_{n,r}$ the distributions are improper (the matrix $\boldsymbol{K}_t$ is singular) we cannot apply any discounting techniques for the specification of $\boldsymbol{W}_t$. Clearly, there is no indication of what the elements of $\boldsymbol{W}_t$ may be. According to [30] or [51, page 133], a practical action is to consider $\boldsymbol{W}_t = \mathbf{O}$ and recalculate the updating forms of $\boldsymbol{H}_t$, $\boldsymbol{h}_t$ of Theorem 4.12. Then, for any $0 < t < t_{n,r}$ we proceed with $\boldsymbol{W}_t = \mathbf{O}$, while for $t \geq t_{n,r}$ when the distributions are proper we use any discounting techniques to specify the non-singular variance $\boldsymbol{W}_t$. In the above references it is stated that no loss happens if we assume $\boldsymbol{W}_t = \mathbf{O}$ when the distributions are improper. This is true because for any $t = 1, \ldots, t_{n,r} - 1$, we have r observations of information and this information will be used to specify the joint state distribution. So we do not aim to detect any changes and so a static model is no loss for the later proper distributions. Of course forecasting at this early stage is not advisable. However, if r is quite large, this procedure may be not desirable, especially when fast response is needed. In such a case a fully specified model with initial values $\boldsymbol{m}_0$, $\boldsymbol{C}_0$, $\boldsymbol{S}_0$, and $\boldsymbol{N}_0$ is recommended. Following we have the version of Theorem 4.12 in the case of $\boldsymbol{W}_t = \mathbf{O}$.

Theorem 4.13. *In the framework of Theorem 4.12, suppose that $\boldsymbol{G}_t$ is non-singular and $\boldsymbol{W}_t = \mathbf{O}$. Then the prior and posterior distributions of Θ_t and*

Σ *have the forms of Theorem 4.12 with recursions defined as follows:*

$$\boldsymbol{H}_t = \boldsymbol{G}_t'^{-1}\boldsymbol{K}_{t-1}\boldsymbol{G}_t^{-1}$$

$$\boldsymbol{h}_t = \boldsymbol{G}_t'^{-1}\boldsymbol{k}_{t-1}$$

$$\Lambda_t = \Delta_{t-1}.$$

Proof. Again the proof is inductive. Suppose first that $p(\Theta_{t-1}, \Sigma|D_{t-1})$ has the stated form. That is

$$\begin{aligned} p(\Theta_{t-1}, \Sigma|D_{t-1}) \;\propto\; & |\Sigma|^{-\frac{1}{2}\left(\frac{\text{trace}(\boldsymbol{\Gamma}_{t-1})}{r}+r+1\right)}\exp\{-\frac{1}{2}\text{trace}\{[\Theta_{t-1}'\boldsymbol{K}_{t-1}\Theta_{t-1} \\ & -2\Theta_{t-1}'\boldsymbol{k}_{t-1} + \Delta_{t-1}]\Sigma^{-1}\}\}. \end{aligned} \qquad (4.26)$$

Then, the system equation is $\Theta_t = \boldsymbol{G}_t\Theta_{t-1}$, which from the non-singularity of $\boldsymbol{G}_t$ becomes $\Theta_{t-1} = \boldsymbol{G}_t^{-1}\Theta_t$ and the Jacobian of Θ_{t-1} with respect to Θ_t, $\mathrm{J}(\Theta_t \to \Theta_{t-1}) = |\boldsymbol{G}_t|^{-r}$, is non zero and constant, (see equation (A.8) in Appendix A.6). So the prior distribution of $(\Theta_t, \Sigma|D_{t-1})$ is obtained from equation (4.26) as

$$\begin{aligned} p(\Theta_t, \Sigma|D_{t-1}) \;\propto\; & |\Sigma|^{-\frac{1}{2}\left(\frac{\text{trace}(\boldsymbol{\Gamma}_{t-1})}{r}+r+1\right)}\exp\{-\frac{1}{2}\text{trace}\{[\Theta_t'\boldsymbol{G}_t'^{-1}\boldsymbol{K}_{t-1} \\ & \times\boldsymbol{G}_t^{-1}\Theta_t - 2\Theta_t'\boldsymbol{G}_t'^{-1}\boldsymbol{k}_{t-1} + \Delta_{t-1}]\Sigma^{-1}\}\}. \end{aligned}$$

Now, multiplying by the likelihood, the posterior $p(\Theta_t, \Sigma|D_t)$ is

$$\begin{aligned} p(\Theta_t, \Sigma|D_t) \;\propto\; & |\Sigma|^{-\frac{1}{2}\left(\frac{\text{trace}(\boldsymbol{\Gamma}_t)}{r}+r+1\right)}\exp\{-\frac{1}{2}\text{trace}\{[\Theta_t'\boldsymbol{K}_t\Theta_t - 2\Theta_t'\boldsymbol{k}_t \\ & +\Delta_t]\Sigma^{-1}\}\}. \end{aligned}$$

Initially for $t = 1$

$$\begin{aligned} p(\Theta_1, \Sigma|D_1) \;\propto\; & |\Sigma|^{-(r+2)/2}\exp\{-\frac{1}{2}\text{trace}\{[\Theta_1'\boldsymbol{F}_1V_1^{-1}\boldsymbol{F}_1'\Theta_1 \\ & +(\boldsymbol{Y}_1 - 2\Theta_1'\boldsymbol{F}_1)V_1^{-1}\boldsymbol{Y}_1']\Sigma^{-1}\}\}, \end{aligned}$$

and the proof follows by induction. □

The next theorems provide the retrospective or filtered distributions.

Theorem 4.14. *In the framework of Theorem 4.12, the filtered distributions of the ECCM for times $t-x$, $x = 0, 1, 2, \ldots, t-1$, are defined as follows*

(1) If $t \le t_{n,r} - 1$, then

$$\begin{aligned} p(\Theta_{t-x}, \Sigma | D_t) \;\; \propto \;\; & |\Sigma|^{-\frac{1}{2}\left(\frac{trace(\Gamma_t)}{r}+r+1\right)} exp\{-\frac{1}{2} trace\{[\Theta'_{t-x} K_t(-x) \\ & \times \Theta_{t-x} - 2\Theta'_{t-x} k_t(-x) + \Delta_t(-x)]\Sigma^{-1}\}\}, \end{aligned}$$

where the defining quantities may be calculated recursively backwards in time according to

$$\begin{aligned} K_t(-x) &= G'_{t-x+1} W^{-1}_{t-x+1} G_{t-x+1} \\ &\quad - G'_{t-x+1} W^{-1}_{t-x+1} P_t^{-1}(-x+1) W^{-1}_{t-x+1} G_{t-x+1} + K_{t-x}, \\ P_t(-x+1) &= W^{-1}_{t-x+1} + K_t(-x+1) - H_{t-x+1}, \\ k_t(-x) &= k_{t-x} + G'_{t-x+1} W^{-1}_{t-x+1} P_t^{-1}(-x+1)[k_t(-x+1) \\ &\quad - h_{t-x+1}], \\ \Delta_t(-x) &= \Delta_t(-x+1) + \Delta_{t-x} - \Lambda_{t-x+1} \\ &\quad - [k_t(-x+1) - h_{t-x+1}]' P_t^{-1}(-x+1)[k_t(-x+1) \\ &\quad - h_{t-x+1}], \end{aligned}$$

and H_t, h_t, K_t, k_t, Λ_t, Δ_t, Γ_t are as defined in Theorem 4.12. Starting values for these recursions are $K_t(0) = K_t$, $k_t(0) = k_t$, and $\Delta_t(0) = \Delta_t$.

(2) If $t \ge t_{n,r}$, then

$$p(\Theta_{t-x}, \Sigma | D_t) \sim NGW^{-1}[a_t(-x), R_t(-x), S_t, N_t, m_t],$$

where

$$\boldsymbol{a}_t(-x) = \boldsymbol{K}_t^{-1}(-x)\boldsymbol{k}_t(-x), \qquad \boldsymbol{R}_t(-x) = \boldsymbol{K}_t^{-1}(-x),$$

$$\boldsymbol{N}_t^{1/2}\boldsymbol{S}_t\boldsymbol{N}_t^{1/2} = \boldsymbol{\Delta}_t(-x) - \boldsymbol{k}_t'(-x)\boldsymbol{a}_t(-x), \qquad \boldsymbol{N}_t = \boldsymbol{\Gamma}_t - (r+n-1)\boldsymbol{I},$$

and as usual $m_t = r + \frac{trace(\boldsymbol{N}_t)}{2r}$.

Proof. The proof is by induction. Assume the theorem to be true for $x-1$. The filtered distributions are defined by

$$p(\Theta_{t-x}, \Sigma|D_t) = \int p(\Theta_{t-x}|\Sigma, \Theta_{t-x+1}, D_t)p(\Theta_{t-x+1}, \Sigma|D_t)\, d\Theta_{t-x+1}, \quad (4.27)$$

for $x = 0, 1, \ldots, t-1$.

Using Bayes' theorem, the first integrand term is

$$p(\Theta_{t-x}|\Sigma, \Theta_{t-x+1}, D_t) = \frac{p(\Theta_{t-x}|\Sigma, \Theta_{t-x+1}, D_{t-x})p(\boldsymbol{Y}|\Theta_{t-x}, \Sigma, \Theta_{t-x+1}, D_{t-x})}{p(\boldsymbol{Y}|\Sigma, \Theta_{t-x+1}, D_{t-x})}, \quad (4.28)$$

where $\boldsymbol{Y} = \{\boldsymbol{Y}_{t-x+1}, \ldots, \boldsymbol{Y}_t\}$. Now given Σ and Θ_{t-x+1}, $\boldsymbol{Y}$ is independent of the previous value Θ_{t-x} so that the two terms $p(\boldsymbol{Y}|\cdot)$ cancel. By Bayes' theorem, the remaining term is

$$p(\Theta_{t-x}|\Theta_{t-x+1}, \Sigma, D_{t-x}) = \frac{p(\Theta_{t-x+1}|\Theta_{t-x}, \Sigma, D_{t-x})p(\Theta_{t-x}|\Sigma, D_{t-x})}{p(\Theta_{t-x+1}|\Sigma, D_{t-x})}.$$

Multiply the nominator and denominator by $p(\Sigma|D_{t-x})$. Then equation (4.28) becomes

$$p(\Theta_{t-x}|\Sigma, \Theta_{t-x+1}, D_t) = \frac{p(\Theta_{t-x+1}|\Theta_{t-x}, \Sigma, D_{t-x})p(\Theta_{t-x}, \Sigma|D_{t-x})}{p(\Theta_{t-x+1}, \Sigma|D_{t-x})}.$$

So equation (4.27) becomes

$$\begin{aligned} p(\Theta_{t-x}, \Sigma|D_t) &= \int \frac{p(\Theta_{t-x+1}|\Theta_{t-x}, \Sigma, D_{t-x})p(\Theta_{t-x}, \Sigma|D_{t-x})}{p(\Theta_{t-x+1}, \Sigma|D_{t-x})} \\ &\quad \times p(\Theta_{t-x+1}, \Sigma|D_t)\, d\Theta_{t-x+1}, \end{aligned}$$

with

$$
\begin{aligned}
(\Theta_{t-x+1}|\Theta_{t-x}, \Sigma, D_{t-x}) &\sim \mathrm{N}[\boldsymbol{G}_{t-x+1}\Theta_{t-x}, \boldsymbol{W}_{t-x+1}, \Sigma], && (4.29)\\
p(\Theta_{t-x}, \Sigma|D_{t-x}) &\propto |\Sigma|^{-\frac{1}{2}(\frac{\mathrm{trace}(\boldsymbol{\Gamma}_{t-x})}{r}+r+1)}\exp\{-\frac{1}{2}\mathrm{trace}\{[\Theta'_{t-x} \\
&\quad \times \boldsymbol{K}_{t-x}\Theta_{t-x} - 2\Theta'_{t-x}\boldsymbol{k}_{t-x} \\
&\quad +\boldsymbol{\Delta}_{t-x}]\Sigma^{-1}\}\}, && (4.30)\\
p(\Theta_{t-x+1}, \Sigma|D_{t-x}) &\propto |\Sigma|^{-\frac{1}{2}(\frac{\mathrm{trace}(\boldsymbol{\Gamma}_{t-x})}{r}+r+1)}\exp\{-\frac{1}{2}\mathrm{trace}\{[\Theta'_{t-x+1} \\
&\quad \times \boldsymbol{H}_{t-x+1}\Theta_{t-x+1} - 2\Theta'_{t-x+1}\boldsymbol{h}_{t-x+1} \\
&\quad +\boldsymbol{\Lambda}_{t-x+1}]\Sigma^{-1}\}\}, && (4.31)\\
p(\Theta_{t-x+1}, \Sigma|D_t) &\propto |\Sigma|^{-\frac{1}{2}(\frac{\mathrm{trace}(\boldsymbol{\Gamma}_{t})}{r}+r+1)}\exp\{-\frac{1}{2}\mathrm{trace}\{[\Theta'_{t-x+1} \\
&\quad \times \boldsymbol{K}_t(-x+1)\Theta_{t-x+1} - 2\Theta'_{t-x+1}\boldsymbol{k}_t(-x+1) \\
&\quad +\boldsymbol{\Delta}_t(-x+1)]\Sigma^{-1}\}\}, && (4.32)
\end{aligned}
$$

where (4.29) is derived from the evolution equation, (4.30), (4.31) from Theorem 4.12, and (4.32) from the hypothesis of induction.

Hence

$$
\begin{aligned}
p(\Theta_{t-x}, \Sigma|D_t) &\propto |\Sigma|^{-\frac{1}{2}(\frac{\mathrm{trace}(\boldsymbol{\Gamma}_{t})}{r}+r+n+1)} \\
&\quad \times \int \exp\{[(\Theta_{t-x+1} - \boldsymbol{\alpha})'\boldsymbol{A}(\Theta_{t-x+1} - \boldsymbol{\alpha}) \\
&\quad +\boldsymbol{B}]\Sigma^{-1}\}\}\, d\Theta_{t-x+1} \\
&= |\Sigma|^{-\frac{1}{2}(\frac{\mathrm{trace}(\boldsymbol{\Gamma}_{t})}{r}+r+1)}\exp\{-\frac{1}{2}\mathrm{trace}\{\boldsymbol{B}\Sigma^{-1}\}\},
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{A} &= \boldsymbol{K}_t(-x+1) - \boldsymbol{H}_{t-x+1} + \boldsymbol{W}_{t-x+1}^{-1} = \boldsymbol{P}_t(-x+1), \\
\boldsymbol{\alpha} &= \boldsymbol{A}^{-1}[\boldsymbol{k}_t(-x+1) - \boldsymbol{h}_{t-x+1} + \boldsymbol{W}_{t-x+1}^{-1}\boldsymbol{G}_{t-x+1}\Theta_{t-x}], \\
\boldsymbol{B} &= \Theta_{t-x}'[\boldsymbol{K}_{t-x} + \boldsymbol{G}_{t-x+1}'\boldsymbol{W}_{t-x+1}^{-1}\boldsymbol{G}_{t-x+1}]\Theta_{t-x} - 2\Theta_{t-x}'\boldsymbol{k}_{t-x} \\
&\quad + \boldsymbol{\Delta}_t(-x+1) + \boldsymbol{\Delta}_{t-x} - \boldsymbol{\Lambda}_{t-x+1} - \boldsymbol{\alpha}'\boldsymbol{A}\boldsymbol{\alpha} \\
&= \Theta_{t-x}'\boldsymbol{K}_t(-x)\Theta_{t-x} - 2\Theta_{t-x}'\boldsymbol{k}_t(-x) + \boldsymbol{\Delta}_t(-x),
\end{aligned}
$$

with $\boldsymbol{K}_t(-x)$, $\boldsymbol{k}_t(-x)$, $\boldsymbol{\Delta}_t(-x)$, as defined in the theorem. The proof of (1) is completed by applying the above analysis with $x = 1$ when $\boldsymbol{K}_t(0) = \boldsymbol{K}_t$, $\boldsymbol{k}_t(0) = \boldsymbol{k}_t$, and $\boldsymbol{\Delta}_t(0) = \boldsymbol{\Delta}_t$.

Now proving (2), if $t \geq t_{n,r}$ the distributions are proper and $p(\Theta_{t-x}, \Sigma | D_t)$ is written as

$$
\begin{aligned}
p(\Theta_{t-x}, \Sigma | D_t) \propto\; & |\Sigma|^{-n/2}\exp\{-\frac{1}{2}\text{trace}\{(\Theta_{t-x} - \boldsymbol{a}_t(-x))'\boldsymbol{R}_t^{-1}(-x)(\Theta_{t-x} \\
& -\boldsymbol{a}_t(-x))\Sigma^{-1}\}\}|\Sigma|^{-\frac{1}{2}\left(\frac{\text{trace}(\boldsymbol{\Gamma}_t)}{r}+r-n+1\right)} \\
& \times\exp\{-\frac{1}{2}\text{trace}\{[\boldsymbol{\Delta}_t(-x) - \boldsymbol{k}_t'(-x)\boldsymbol{a}_t(-x)]\Sigma^{-1}\}\},
\end{aligned}
$$

where

$$
\boldsymbol{a}_t(-x) = \boldsymbol{K}_t^{-1}(-x)\boldsymbol{k}_t(-x) \qquad \text{and} \qquad \boldsymbol{R}_t(-x) = \boldsymbol{K}_t^{-1}(-x).
$$

From the matrix normal component

$$
(\Theta_{t-x} | \Sigma, D_t) \sim \mathrm{N}[\boldsymbol{a}_t(-x), \boldsymbol{R}_t(-x), \Sigma],
$$

and from the generalized inverse Wishart component

$$
(\Sigma | D_t) \sim \mathrm{GW}^{-1}[\boldsymbol{S}_t, \boldsymbol{N}_t, m_t],
$$

where

$$N_t^{1/2} S_t N_t^{1/2} = \Delta_t(-x) - k_t'(-x) a_t(-x) \qquad \text{and} \qquad N_t = \Gamma_t - (r+n-1)I.$$

Thus, the distribution $p(\Theta_{t-x}, \Sigma | D_t)$ is obtained immediately. □

The next theorem provides the filtered distributions in the case of $W_t = O$.

Theorem 4.15. *In the framework of Theorem 4.14, suppose that $W_t = O$. Then the filtered distributions are as in Theorem 4.14 with the following changes to the recursion:*

$$K_t(-x) = G'_{t-x+1} K_t(-x+1) G_{t-x+1},$$

$$k_t(-x) = G'_{t-x+1} k_t(-x+1),$$

$$\Delta_t(-x) = \Delta_t(-x+1).$$

Proof. Again the proof is inductive. Suppose that $p(\Theta_{t-x+1}, \Sigma | D_t)$ has the stated form. That is

$$\begin{aligned} p(\Theta_{t-x+1}, \Sigma | D_t) \;\propto\; & |\Sigma|^{-\frac{1}{2}\left(\frac{\text{trace}(\Gamma_t)}{r}+r+1\right)} \exp\{-\frac{1}{2}\text{trace}\{[\Theta'_{t-x+1} \\ & \times K_t(-x+1)\Theta_{t-x+1} - 2\Theta'_{t-x+1} k_t(-x+1) \\ & + \Delta_t(-x+1)]\Sigma^{-1}\}\}. \end{aligned} \tag{4.33}$$

Then, the system equation is $\Theta_{t-x+1} = G_{t-x+1}\Theta_{t-x}$ and substitution to equation (4.33) provides the distribution $p(\Theta_{t-x}, \Sigma | D_t)$, with $K_t(-x)$, $k_t(-x)$, and $\Delta_t(-x)$ as defined in the theorem. The starting values of Theorem 4.14 complete the proof. □

Note that by contrast with Theorem 4.13, Theorem 4.15 does not require the non-singularity of G_t.

4.7 An Approximation of the GMDLM

The normal GMDLM is defined as

$$\boldsymbol{Y}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim \mathrm{N}[\mathbf{0}, \Sigma], \tag{4.34}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathrm{N}[\mathbf{0}, \boldsymbol{W}_t], \tag{4.34$'$}$$

where $\dim(\boldsymbol{Y}_t) = r \times 1$, $\dim(\boldsymbol{F}_t) = n \times r$, $\dim(\boldsymbol{\theta}_t) = n \times 1$, $\dim(\boldsymbol{\nu}_t) = r \times 1$, $\dim(\Sigma) = r \times r$, $\dim(\boldsymbol{G}_t) = n \times n$, $\dim(\boldsymbol{\omega}_t) = n \times 1$, and $\dim(\boldsymbol{W}_t) = n \times n$. It is also assumed that

$$(\boldsymbol{\theta}_0|D_0) \sim \mathrm{N}[\boldsymbol{m}_0, \boldsymbol{C}_0], \tag{4.35}$$

$$\Sigma \sim \mathrm{GW}^{-1}[\boldsymbol{S}_0, \boldsymbol{N}_0, m_0], \tag{4.36}$$

for some known quantities $\boldsymbol{m}_0$, $\boldsymbol{C}_0$, $\boldsymbol{S}_0$, and $\boldsymbol{N}_0$.

However, it is evident that as the matrices Σ, $\boldsymbol{W}_t$ are of different dimensions, the usual scaling cannot provide a tractable analysis. Even in the special case of $n = r$ conjugacy is not possible unless matrices $\boldsymbol{G}_t$, $\Sigma^{1/2}$ and $\boldsymbol{F}_t'$, $\Sigma^{1/2}$ are commutative, setting $\boldsymbol{W}_t = \Sigma^{1/2}\boldsymbol{W}_t^*\Sigma^{1/2}$ for a known variance matrix $\boldsymbol{W}_t^*$. Barbosa and Harrison ([4]), proposed an approximation, which is exact for the univariate case and the CCM. They found that the method is significantly faster and more reliable than standard approximations (Student t filter, robust filter, see [49]). However, they encountered problems with missing observations and expert intervention. That is partially due to lack of the unconditional posterior distribution of $(\boldsymbol{\theta}_t|D_t)$, the method only provides the distribution of $(\boldsymbol{\theta}_t|D_t, \Sigma = \boldsymbol{S}_0)$, (a problem considered in [4]) and partially due to the explicit use of the inverse Wishart distribution. To the following we present an upgraded version of the main theorem that will allow flexible intervention as well as missing observation analysis.

Definition 4.4. *A scaled version of the GMDLM is defined by equations (4.34), (4.34′) with the following distributional assumptions*

(i) $(\boldsymbol{Y}_t|\boldsymbol{\mu}_t, \Sigma) \sim N[\boldsymbol{\mu}_t, \Sigma]$,

(ii) $(\boldsymbol{\mu}_t|\Sigma, D_{t-1}) \sim N[\boldsymbol{f}_t, \boldsymbol{R}_t^*]$,

(iii) $(\Sigma|D_{t-1}) \sim GW^{-1}[\boldsymbol{S}_{t-1}, \boldsymbol{N}_{t-1}, m_{t-1}]$,

where $\boldsymbol{\mu}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t$ *is the mean response vector,* $\boldsymbol{f}_t$ *is the prior mean for* $\boldsymbol{\mu}_t$ *given* Σ*, and* $\boldsymbol{R}_t^* = \Sigma^{1/2}\boldsymbol{X}_t\Sigma^{1/2}$ *is the prior variance matrix for* $\boldsymbol{\mu}_t$ *given* Σ*, where* $\boldsymbol{X}_t = V[\boldsymbol{\mu}_t|\Sigma = \boldsymbol{S}_0, D_{t-1}]$ *for a known starting matrix* $\boldsymbol{S}_0$.

Further we state the approximations, necessary for the analysis that are essentially the same as in [4].

$$\text{If} \quad (\boldsymbol{\mu}_t|\Sigma = \boldsymbol{S}_0, D_{t-1}) \sim \mathrm{N}[\boldsymbol{f}_t, \boldsymbol{S}_0^{1/2}\boldsymbol{X}_t\boldsymbol{S}_0^{1/2}] \quad \text{then}$$

$$(\boldsymbol{\mu}_t|\Sigma, D_{t-1}) \sim \mathrm{N}[\boldsymbol{f}_t, \Sigma^{1/2}\boldsymbol{X}_t\Sigma^{1/2}], \tag{4.37}$$

$$\Sigma^{1/2}\boldsymbol{Q}_t^*\Sigma^{1/2} \approx \boldsymbol{S}_{t-1}^{1/2}\boldsymbol{Q}_t^*\boldsymbol{S}_{t-1}^{1/2}, \tag{4.38}$$

$$\boldsymbol{S}_{t-1} \approx \Sigma, \tag{4.39}$$

$$\Sigma^{1/2}\boldsymbol{A}_t^1\Sigma^{-1/2} \approx \boldsymbol{S}_t^{1/2}\boldsymbol{A}_t^1\boldsymbol{S}_t^{-1/2}, \tag{4.40}$$

for some known quantities $\boldsymbol{Q}_t$, $\boldsymbol{Q}_t^*$, and $\boldsymbol{A}_t^1$.

Note that assumptions (4.38), (4.40) are essentially embedded into assumption (4.39). However, we chose the above writing so that access and reference of these assumptions to the following theorem are straight-forward.

Theorem 4.16. *Under assumptions (4.37), (4.38), (4.39), and (4.40) for the scaled DLM the next approximate distributional results apply*

(a) *Posterior distributions at* $t-1$*:*

$$(\Sigma|D_{t-1}) \sim GW^{-1}[\boldsymbol{S}_{t-1}, \boldsymbol{N}_{t-1}, m_{t-1}],$$

$$(\boldsymbol{\theta}_{t-1}|\Sigma = \boldsymbol{S}_0, D_{t-1}) \sim N[\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}],$$

for some known quantities $\boldsymbol{m}_{t-1}$, $\boldsymbol{C}_{t-1}$, $\boldsymbol{S}_{t-1}$, $\boldsymbol{N}_{t-1}$.

(b) *Prior distribution at* t*:*

$$(\boldsymbol{\theta}_t|\Sigma = \boldsymbol{S}_0, D_{t-1}) \sim N[\boldsymbol{a}_t, \boldsymbol{R}_t],$$

where

$$\boldsymbol{a}_t = \boldsymbol{G}_t \boldsymbol{m}_{t-1} \qquad \text{and} \qquad \boldsymbol{R}_t = \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t' + \boldsymbol{W}_t.$$

(c) *One-step forecast distribution at* t*:*

$$(\boldsymbol{Y}_t|\Sigma, D_{t-1}) \sim N[\boldsymbol{f}_t, \boldsymbol{Q}_t],$$

where

$$\boldsymbol{f}_t = \boldsymbol{F}_t' \boldsymbol{a}_t \qquad \text{and} \qquad \boldsymbol{Q}_t = \boldsymbol{S}_{t-1}^{1/2} \boldsymbol{Q}_t^* \boldsymbol{S}_{t-1}^{1/2},$$

with

$$\boldsymbol{Q}_t^* = \boldsymbol{I} + \boldsymbol{X}_t \qquad \text{and} \qquad \boldsymbol{X}_t = \boldsymbol{F}_t' \boldsymbol{R}_t \boldsymbol{F}_t.$$

(d) *Posterior distributions at* t*:*

$$(\Sigma|D_t) \sim GW^{-1}[\boldsymbol{S}_t, \boldsymbol{N}_t, m_t],$$

$$(\boldsymbol{\theta}_t|\Sigma = \boldsymbol{S}_0, D_t) \sim N[\boldsymbol{m}_t, \boldsymbol{C}_t],$$

where

$$\boldsymbol{N}_t^{1/2} \boldsymbol{S}_t \boldsymbol{N}_t^{1/2} = \boldsymbol{N}_{t-1}^{1/2} \boldsymbol{S}_{t-1} \boldsymbol{N}_{t-1}^{1/2} + \boldsymbol{h}_t \boldsymbol{h}_t' \qquad \text{and} \qquad \boldsymbol{N}_t = \boldsymbol{N}_{t-1} + \boldsymbol{I},$$

with

$$h_t = S_{t-1}^{1/2} Q_t^{-1/2} e_t, \qquad e_t = Y_t - f_t,$$

$$m_t = a_t + A_t^* A_t e_t, \qquad C_t = R_t - A_t^* S_0^{1/2} A_t Q_t^* A_t' S_0^{1/2} A_t^{*\prime},$$

and

$$A_t^* = R_t F_t (S_0^{1/2} X_t S_0^{1/2})^{-1}, \quad A_t = S_t^{1/2} A_t^1 S_t^{-1/2}, \quad A_t^1 = X_t Q_t^{*-1}.$$

Proof. The proof is by induction. Suppose (a) be true. (b) follows immediately, simply being the case of known variances. Now denote $\mu_t = F_t'\theta_t$ as usual and notice that $(\mu_t|\Sigma = S_0, D_{t-1}) \sim \mathrm{N}[f_t, S_0^{1/2} X_t S_0^{1/2}]$, for f_t, X_t as stated in the theorem. Thus, approximation (4.37) yields $(\mu_t|\Sigma, D_{t-1}) \sim \mathrm{N}[f_t, \Sigma^{1/2} X_t \Sigma^{1/2}]$, which together with the observation equation gives

$$(Y_t|\Sigma, D_{t-1}) \sim \mathrm{N}[f_t, \Sigma^{1/2} Q_t^* \Sigma^{1/2}].$$

Now applying approximation (4.38), we have (c). It follows that

$$\begin{pmatrix} \mu_t \\ Y_t \end{pmatrix} \Bigg| \Sigma, D_{t-1} \Bigg) \sim \mathrm{N}\left[\begin{pmatrix} f_t \\ f_t \end{pmatrix}, \begin{pmatrix} \Sigma^{1/2} X_t \Sigma^{1/2} & \Sigma^{1/2} X_t \Sigma^{1/2} \\ \Sigma^{1/2} X_t \Sigma^{1/2} & \Sigma^{1/2} Q_t^* \Sigma^{1/2} \end{pmatrix}\right].$$

Now form the joint distribution of θ_t, μ_t, Y_t given $\Sigma = S_0$ and D_{t-1}

$$\left(\begin{matrix} \theta_t \\ \mu_t \\ Y_t \end{matrix} \,\middle|\, \Sigma = S_0, D_{t-1}\right) \sim \mathrm{N}\left[\begin{pmatrix} a_t \\ f_t \\ f_t \end{pmatrix}, \begin{pmatrix} R_t & R_t F_t & R_t F_t \\ F_t' R_t & S_0^{1/2} X_t S_0^{1/2} & S_0^{1/2} X_t S_0^{1/2} \\ F_t' R_t & S_0^{1/2} X_t S_0^{1/2} & S_0^{1/2} Q_t^* S_0^{1/2} \end{pmatrix}\right].$$

Defining Q_t as in the theorem and $e_t = Y_t - f_t$, $(e_t|\Sigma, D_{t-1}) \sim \mathrm{N}[0, \Sigma^{1/2} Q_t^* \Sigma^{1/2}]$, so $(\Sigma^{1/2} Q_t^{-1/2} S_{t-1}^{1/2} \Sigma^{-1/2} e_t|\Sigma, D_{t-1}) \sim \mathrm{N}[0, \Sigma]$ and using approximation (4.39)

$$(S_{t-1}^{1/2} Q_t^{-1/2} e_t|\Sigma, D_{t-1}) \sim \mathrm{N}[0, \Sigma].$$

This equation in conjunction with $(\Sigma|D_{t-1}) \sim \mathrm{GW}^{-1}[\boldsymbol{S}_{t-1}, \boldsymbol{N}_{t-1}, m_{t-1}]$ gives

$$(\Sigma|D_t) \sim \mathrm{GW}^{-1}[\boldsymbol{S}_t, \boldsymbol{N}_t, m_t],$$

where

$$\boldsymbol{N}_t^{1/2}\boldsymbol{S}_t\boldsymbol{N}_t^{1/2} = \boldsymbol{N}_{t-1}^{1/2}\boldsymbol{S}_{t-1}\boldsymbol{N}_{t-1}^{1/2} + \boldsymbol{S}_{t-1}^{1/2}\boldsymbol{Q}_t^{-1/2}\boldsymbol{e}_t\boldsymbol{e}_t'\boldsymbol{Q}_t^{-1/2}\boldsymbol{S}_{t-1}^{1/2}$$

and

$$\boldsymbol{N}_t = \boldsymbol{N}_{t-1} + \boldsymbol{I},$$

which is essentially the updating of (d). Now define $\boldsymbol{A}_t^1 = \boldsymbol{X}_t\boldsymbol{Q}_t^{*-1}$ and $\boldsymbol{A}_t = \boldsymbol{S}_t^{1/2}\boldsymbol{A}_t^1\boldsymbol{S}_t^{-1/2}$. According to standard conditional results, the posterior distribution of $\boldsymbol{\mu}_t$ is obtainable from

$$(\boldsymbol{\mu}_t|\Sigma, D_t) \sim \mathrm{N}[\boldsymbol{f}_t + \Sigma^{1/2}\boldsymbol{A}_t^1\Sigma^{-1/2}\boldsymbol{e}_t, \Sigma^{1/2}(\boldsymbol{X}_t - \boldsymbol{A}_t^1\boldsymbol{Q}_t^*\boldsymbol{A}_t^{1'})\Sigma^{1/2}],$$

using approximation (4.40) as

$$(\boldsymbol{\mu}_t|\Sigma, D_t) \sim \mathrm{N}[\boldsymbol{f}_t + \boldsymbol{A}_t\boldsymbol{e}_t, \Sigma^{1/2}(\boldsymbol{X}_t - \boldsymbol{A}_t^1\boldsymbol{Q}_t^*\boldsymbol{A}_t^{1'})\Sigma^{1/2}]. \tag{4.41}$$

Writing $\boldsymbol{A}_t^* = \boldsymbol{R}_t\boldsymbol{F}_t(\boldsymbol{S}_0^{1/2}\boldsymbol{X}_t\boldsymbol{S}_0^{1/2})^{-1}$ and considering the conditional independence of $\boldsymbol{\theta}_t$ and $\boldsymbol{Y}_t$ given $\boldsymbol{\mu}_t$ from equation (4.41) and the prior $(\boldsymbol{\theta}_t|\Sigma = \boldsymbol{S}_0, D_{t-1})$ we obtain

$$(\boldsymbol{\theta}_t|\Sigma = \boldsymbol{S}_0, D_t) \sim \mathrm{N}[\boldsymbol{m}_t, \boldsymbol{C}_t],$$

with moments $\boldsymbol{m}_t$, $\boldsymbol{C}_t$ as stated in the theorem. The proof is completed by noting that (a) is true for $t = 1$. □

The above theorem implies that this approximation is exact for the ECCM, while it is expected to be good for the general multivariate DLM. Missing observation analysis of this model is considered in Chapter 6.

CHAPTER 5

The General Multivariate DLM

5.1 Introduction

In this chapter the general linear multivariate, but not necessarily normal, DLM is considered. A weak condition, which is exact for the normal model, allows a very flexible and widely used analysis. This follows [42]. Section 5.2 introduces the general problem of probability modelling and addresses the need for consistent probability modelling procedures. The conceptual basis for this is taken by some unpublished work by P.J. Harrison. In Section 5.2.2 the weak prior posterior probability modelling (see [51, chapter 4]) is extended to the multivariate case with unknown observational variance matrices. In Section 5.3 the estimate of the unknown variance matrix is considered and all the necessary material for the DLM analysis is recollected.

The following section develops the main DLM procedure, the k-step ahead forecast distributions, and the filtered distributions. In Section 5.5 some limiting results are presented. The last section of the chapter is a simulation showing the capabilities of the underlying methods.

5.2 Probability Modelling

5.2.1 The General Problem

In Chapters 3 and 4 probability modelling is based on regression and more general Bayesian methods. These methods are formally consistent in the sense that their assumptions are consistent with some probability space or can be derived from a probability model.

A full probability model for a random vector $\boldsymbol{Z}$ comprises the joint distribution of all its elements. The forecast of any function of $\boldsymbol{Z}$ is then just that function's marginal distribution. For example let $\boldsymbol{Z}' = (\boldsymbol{X}', \boldsymbol{Y}')$ and let the density of $\boldsymbol{Z}$ be denoted by $p(\cdot)$. The forecast of $\boldsymbol{X}$ is

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y}.$$

When the value of $\boldsymbol{Y} = \boldsymbol{y}^*$ is observed the revised probability model for $\boldsymbol{Z}$, or for its unknown part $\boldsymbol{X}$, since $\boldsymbol{Y}$ is known to be $\boldsymbol{y}^*$, is

$$p(\boldsymbol{x}|\boldsymbol{Y} = \boldsymbol{y}^*) = \frac{p(\boldsymbol{x}, \boldsymbol{y}^*)}{p(\boldsymbol{y}^*)} = \frac{p(\boldsymbol{Y} = \boldsymbol{y}^*|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{Y} = \boldsymbol{y}^*)},$$

from direct application of the Bayes' theorem.

The general forecasting situation is that we are going to receive some information $\boldsymbol{Y}$, which relates to $\boldsymbol{X}$ and that we are going to use it to produce a forecast for $\boldsymbol{X}$. The full probability model approach is to construct a joint probability distribution for $(\boldsymbol{X}, \boldsymbol{Y})$, or alternatively for $(\boldsymbol{X}, \boldsymbol{U})$ where, given $\boldsymbol{Y}$, $\boldsymbol{U}$ is sufficient for $\boldsymbol{X}$ so that

$$\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} | \boldsymbol{U}.$$

Then, upon observing $\boldsymbol{Y} = \boldsymbol{y}^*$, the corresponding conditional distribution $(\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y}^*) \equiv (\boldsymbol{X}|\boldsymbol{U} = \boldsymbol{u}^*)$, is taken using Bayes' theorem as before.

Although the above methodology is theoretically consistent there are problems in the application. Unless normal distributions are assumed, the marginal distribution derivation will require numerical integration, which is not often desirable, especially when retropsective analysis is needed. Moreover, given a set of data it is not always easy to fit a distribution and again approximate techniques like smoothing may not be desirable. This problem may be resolved by relaxing the mutual independence assumptions and providing a modelling approach that leaves the distributions unspecified. The modeller does not need to specify any complicated distribution. However, the forecast distribution is not fully determined. Only the mean and variance are specified, which is usually more than satisfactory since in most practical situations this is enough. Note that the benefit in the multivariate case is greater, since the distributions are rather complex and often there is not much justification for the choice of distributions of the error terms. The errors are traditionally modelled with a normal distribution, but this may lead to inaccurate forecasts.

5.2.2 Weak Posterior Prior Probability Modelling

The following development of the concept of *Weak Probability* as well as its application to the DLM theory follow [42].

Definition 5.1. *Suppose X, Y, U are any random vectors with a joint distribution. Then given U, X is said to be second order independent of Y, written $X \perp_2 Y|U$, if and only if*

$$E[X|U,Y] = E[X|U],$$
$$E[XX'|U,Y] = E[XX'|U],$$
$$E[XY'|U] = E[X|U]E[Y'|U].$$

X and Y are said to be second order mutually independent given U, writing $X \perp\!\!\!\perp_2 Y|U$, if and only if

$$X \perp_2 Y|U \qquad \textit{and} \qquad Y \perp_2 X|U.$$

Similarly the first order independence and first order mutual independence can be defined, just excluding all U's from definition 5.1, written $X \perp_1 Y$ and $X \perp\!\!\!\perp_1 Y$ respectively.

It follows that if $X(\mathcal{U})Y|U$ denotes that X is uncorrelated with Y, given U, then

$$X \perp\!\!\!\perp Y|U \Rightarrow X \perp\!\!\!\perp_2 Y|U \Rightarrow X \perp_2 Y|U \Rightarrow X(\mathcal{U})Y|U.$$

Definition 5.2. *A first order Posterior Prior model, PP(1) is defined such that, (a) $\phi(X,Y)$ is mutually independent of Y and (b) the distribution of $(X|Y = y)$ is obtainable from the distribution of $\phi(X,Y = y)$, for some suitable function $\phi(\cdot)$ of the vectors X, Y. The notation is $(X,Y;\phi)$ is PP(1).*

Example 1: Suppose that the vectors X, Y have a joint normal distribution such that

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} m_x \\ m_y \end{pmatrix}, \begin{pmatrix} V_x & A_{xy}V_y \\ V_yA'_{xy} & V_y \end{pmatrix}\right],$$

for some known moments $\boldsymbol{m}_x$, $\boldsymbol{m}_y$, $\boldsymbol{V}_x$, $\boldsymbol{V}_y$, where $\boldsymbol{A}_{xy}$ is the regression matrix of $\boldsymbol{X}$ on $\boldsymbol{Y}$.

It is known that $\boldsymbol{X} - \boldsymbol{A}_{xy}\boldsymbol{Y}$ is independent of $\boldsymbol{Y}$ and

$$\begin{pmatrix} \boldsymbol{X} - \boldsymbol{A}_{xy}\boldsymbol{Y} \\ \boldsymbol{Y} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \boldsymbol{m}_x - \boldsymbol{A}_{xy}\boldsymbol{m}_y \\ \boldsymbol{m}_y \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_x - \boldsymbol{A}_{xy}\boldsymbol{V}_y\boldsymbol{A}'_{xy} & \mathrm{O} \\ \mathrm{O} & \boldsymbol{V}_y \end{pmatrix}\right], \tag{5.1}$$

The conditional distribution of $(\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y})$ is

$$(\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y}) \sim \mathrm{N}[\boldsymbol{m}_{x|y}, \boldsymbol{V}_{x|y}], \tag{5.2}$$

where

$$\boldsymbol{m}_{x|y} = \boldsymbol{m}_x + \boldsymbol{A}_{xy}(\boldsymbol{y} - \boldsymbol{m}_y)$$
$$\boldsymbol{V}_{x|y} = \boldsymbol{V}_x - \boldsymbol{A}_{xy}\boldsymbol{V}_y\boldsymbol{A}'_{xy},$$

for any known value $\boldsymbol{y}$.

The distribution (5.2) is fully known if and only if its moments $\boldsymbol{m}_{x|y}$, $\boldsymbol{V}_{x|y}$ are known. Since these moments are directly calculable from equation (5.1) the distribution of $(\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y})$ is obtained by the distribution of $\boldsymbol{X} - \boldsymbol{A}_{xy}\boldsymbol{Y}$.

So choosing $\phi(\boldsymbol{X}, \boldsymbol{Y}) = \boldsymbol{X} - \boldsymbol{A}_{xy}\boldsymbol{Y}$ the conditions (a), (b) of Definition 5.2 are satisfied, and so $(\boldsymbol{X}, \boldsymbol{Y}; \phi)$ is PP(1).

The next definition is a generalization of Definition 5.2.

Definition 5.3. *Let $\boldsymbol{X}, \boldsymbol{V}, \boldsymbol{Y}$ be random vectors. Then a second order Posterior Prior probability model, PP(2) is such that, $(\boldsymbol{X}, \boldsymbol{Y}|\boldsymbol{V}; \phi)$, $(\boldsymbol{V}, \boldsymbol{Y}; \psi)$ are both PP(1) probability models, for some functions ϕ, ψ. The notation is $(\boldsymbol{X}, \boldsymbol{Y}|\boldsymbol{V}; \phi, \psi)$ is PP(2).*

That is for each value $\boldsymbol{y}$ of $\boldsymbol{Y}$ with $p(\boldsymbol{y}) > 0$,

$$\phi(\boldsymbol{X}, \boldsymbol{Y}) \perp\!\!\!\perp \boldsymbol{Y}|\boldsymbol{V} \quad \text{and} \quad \psi(\boldsymbol{V}, \boldsymbol{Y}) \perp\!\!\!\perp \boldsymbol{Y}$$

and the conditional distribution of $(\boldsymbol{X}|\boldsymbol{y},\boldsymbol{V})$ is obtainable from $\phi(\boldsymbol{X},\boldsymbol{y})$ so that, for all $\boldsymbol{y}$, the distribution of $(\phi(\boldsymbol{X},\boldsymbol{y})|\boldsymbol{V})$ is the prior distribution of $(\phi(\boldsymbol{X},\boldsymbol{Y})|\boldsymbol{V})$ and the distribution of $\psi(\boldsymbol{V},\boldsymbol{y})$ is the prior distribution of $\psi(\boldsymbol{V},\boldsymbol{Y})$.

Example 2: Suppose for ease that $\boldsymbol{X} = X$, $\boldsymbol{Y} = Y$, $\boldsymbol{V} = V$ are scalars such that

$$\left(\begin{matrix} X \\ Y \end{matrix} \middle| V \right) \sim \mathrm{N}\left[\left(\begin{matrix} 0 \\ 0 \end{matrix} \right), V \left(\begin{matrix} 2 & 1 \\ 1 & 1 \end{matrix} \right) \right], \qquad 6V^{-1} \sim \chi^2_6, \qquad V > 0,$$

from which we get

$$(X - Y|V) \sim \mathrm{N}[0, V] \qquad \text{and} \qquad (6 + Y^2)V^{-1} \sim \chi^2_7,$$

where χ^2_n $(n = 6, 7)$ denotes the chi-squared distribution with n degrees of freedom.

By using the Bayes' theorem for any value y of Y, we have

$$p((6 + Y^2)V^{-1}|y) = p((6 + Y^2)V^{-1}),$$

which justifies that $(6 + Y^2)V^{-1} \perp\!\!\!\perp Y$.

Thus, by defining

$$\phi(X, Y) = X - Y \qquad \text{and} \qquad \psi(V, Y) = (6 + Y^2)V^{-1},$$

we have that $(X, Y|V; \phi)$, $(V, Y; \psi)$ are PP(1) models, hence $(X, Y|V; \phi, \psi)$ is PP(2) model.

Now the definition of weak probability models is possible.

Definition 5.4. *The definition of a second order Weak Posterior Prior probability model, WPP(2) follows that of a PP(2) model except that strict conditional independence is replaced by second order independence and the word* **distribution** *is replaced by the* **mean** *and* **variance**.

Similarly WPP(1) models can be defined. Definition 5.4 has a key role to the following analysis. It allows us to keep unspecified the distributions of the model and hence it allows a wider application. To the following we are going to use WPP(2) models, that is for $(\boldsymbol{X}', \boldsymbol{Y}')$ it is $\phi(\boldsymbol{X}, \boldsymbol{Y})$ so that $\mathrm{E}[\phi(\boldsymbol{X}, \boldsymbol{Y} = \boldsymbol{y})] = \mathrm{E}[\phi(\boldsymbol{X}, \boldsymbol{Y})]$, $\mathrm{V}[\phi(\boldsymbol{X}, \boldsymbol{Y} = \boldsymbol{y})] = \mathrm{V}[\phi(\boldsymbol{X}, \boldsymbol{Y})]$, and $\mathrm{C}[\phi(\boldsymbol{X}, \boldsymbol{Y}), \boldsymbol{Y}] = \mathrm{O}$, and is such that the mean and variance of $(\boldsymbol{X} | \boldsymbol{Y} = \boldsymbol{y})$ are obtainable from those of $\phi(\boldsymbol{X}, \boldsymbol{Y} = \boldsymbol{y})$ which are the prior mean and variance of $\phi(\boldsymbol{X}, \boldsymbol{Y})$.

Now we work with the function ϕ. First consider the case of the WPP(1) model.

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be any random vectors with dimensions $n \times 1$, $r \times 1$ respectively and with a joint distribution. Write $\boldsymbol{Z}' = (\boldsymbol{X}', \boldsymbol{Y}')$ such that $\boldsymbol{Z} \sim [\boldsymbol{m}, \boldsymbol{V}]$, with

$$\boldsymbol{m} = \begin{pmatrix} \boldsymbol{m}_x \\ \boldsymbol{m}_y \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_x & \boldsymbol{A}_{xy}\boldsymbol{V}_y \\ \boldsymbol{V}_y\boldsymbol{A}'_{xy} & \boldsymbol{V}_y \end{pmatrix},$$

where $\boldsymbol{m}_x = \mathrm{E}[\boldsymbol{X}]$, $\boldsymbol{m}_y = \mathrm{E}[\boldsymbol{Y}]$, $\boldsymbol{V}_x = \mathrm{V}[\boldsymbol{X}]$, $\boldsymbol{V}_y = \mathrm{V}[\boldsymbol{Y}]$, and $\boldsymbol{A}_{xy}$ the regression matrix of $\boldsymbol{X}$ on $\boldsymbol{Y}$.

Define the $(n+r) \times (n+r)$ matrix $\boldsymbol{L}$ as

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{I}_n & -\boldsymbol{A}_{xy} \\ \mathrm{O} & \boldsymbol{I}_r \end{pmatrix}$$

such that $\boldsymbol{LZ} \sim [\boldsymbol{Lm}, \boldsymbol{LVL}']$. This proves that $\boldsymbol{X} - \boldsymbol{A}_{xy}\boldsymbol{Y}$ is uncorrelated with $\boldsymbol{Y}$.

Now, in order to make inferences about the weak posterior we further assume

$$\boldsymbol{X} - \boldsymbol{A}_{xy}\boldsymbol{Y} \perp_1 \boldsymbol{Y}. \tag{5.3}$$

Assumption (5.3) is correct if a joint normal distribution is used. It is expected that it will be satisfied with approximate normal distributions, e.g. T distributions with suitable degrees of freedom. If WPP(2) models are considered assumption (5.3) is replaced by

$$\boldsymbol{X} - \boldsymbol{A}_{xy}\boldsymbol{Y} \perp_2 \boldsymbol{Y}|\boldsymbol{U}. \tag{5.4}$$

The next theorem is the well known theorem of probability theory and is a key result for the following.

Theorem 5.1. *If $\boldsymbol{X}$, $\boldsymbol{W}$, $\boldsymbol{U}$, and $\boldsymbol{Y}$ are any four random vectors with a joint distribution then*

$$E[\boldsymbol{X}|\boldsymbol{Y}] = E[E(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{Y})|\boldsymbol{Y}],$$
$$V[\boldsymbol{X}|\boldsymbol{Y}] = E[V(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{Y})|\boldsymbol{Y}] + V[E(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{Y})|\boldsymbol{Y}],$$
$$C[\boldsymbol{X},\boldsymbol{W}|\boldsymbol{Y}] = E[C(\boldsymbol{X},\boldsymbol{W}|\boldsymbol{U},\boldsymbol{Y})|\boldsymbol{Y}] + C[E(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{Y}),E(\boldsymbol{W}|\boldsymbol{U},\boldsymbol{Y})|\boldsymbol{Y}].$$

5.3 The Estimate of the Observational Variance Matrix

Consider the following multivariate DLM

$$\boldsymbol{Y}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim [\boldsymbol{0},\boldsymbol{\Sigma}], \tag{5.5}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim [\boldsymbol{0},\boldsymbol{W}_t], \tag{5.5$'$}$$

where $\boldsymbol{Y}_t$ is an $r\times 1$ observation vector, $\boldsymbol{F}_t$ an $n\times r$ design matrix, $\boldsymbol{\theta}_t$ an $n\times 1$ parameter vector, $\boldsymbol{\nu}_t$ an $r\times 1$ random vector, $\boldsymbol{\Sigma}$ an unknown $r\times r$ observation variance matrix, $\boldsymbol{G}_t$ an $n\times n$ evolution matrix, $\boldsymbol{\omega}_t$ an $n\times 1$ random vector, and $\boldsymbol{W}_t$ a known $n \times n$ evolution variance matrix. Note that no distribution of $\boldsymbol{\nu}_t$, $\boldsymbol{\omega}_t$ has been specified. As usual $\boldsymbol{\nu}_t$ and $\boldsymbol{\omega}_t$ are assumed mutually and internally independent.

Let $\boldsymbol{S}_t$ be the estimate of $\boldsymbol{\Sigma}$, given D_t. In order to calculate $\boldsymbol{S}_t$, we introduce the following assumptions

$$\mathrm{vech}(\boldsymbol{\Sigma} - \mathcal{A}_t \boldsymbol{e}_t \boldsymbol{e}_t' \mathcal{A}_t') \perp_1 \boldsymbol{Y}_t | D_{t-1}, \tag{5.6}$$

$$\mathrm{C}[\mathrm{vech}(\boldsymbol{\Sigma}), \mathrm{vech}(\mathcal{A}_t \boldsymbol{e}_t \boldsymbol{e}_t' \mathcal{A}_t') | D_{t-1}] = \mathrm{V}[\mathrm{vech}(\mathcal{A}_t \boldsymbol{e}_t \boldsymbol{e}_t' \mathcal{A}_t') | D_{t-1}], \tag{5.7}$$

where $\mathcal{A}_t = n_t^{-1/2} \boldsymbol{S}_{t-1}^{1/2} \boldsymbol{Q}_t^{-1/2}$, $\boldsymbol{Q}_t = \boldsymbol{F}_t' \boldsymbol{R}_t \boldsymbol{F}_t + \boldsymbol{S}_{t-1}$ is the one-step forecast variance, n_t the degrees of freedom at t, and $\boldsymbol{S}_{t-1}^{1/2}$ and $\boldsymbol{Q}_t^{-1/2}$, denote the square roots of the matrices $\boldsymbol{S}_{t-1}$ and $\boldsymbol{Q}_t^{-1}$ respectively, based on the eigenstructure of the matrices (see Appendix A.7 for details).

We will also need to evaluate $\mathrm{V}[\mathrm{vech}(\boldsymbol{e}_t \boldsymbol{e}_t') | D_{t-1}]$. In order to do this we include the following equation in the model (5.5), (5.5′)

$$\mathrm{vech}(\boldsymbol{e}_t \boldsymbol{e}_t') = \boldsymbol{E}_t \boldsymbol{e}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim [\mathrm{vech}(\boldsymbol{Q}_t), \boldsymbol{V}_{\epsilon,t}], \tag{5.8}$$

where $\boldsymbol{E}_t$ is a known $(r(r+1)/2) \times r$ design matrix and the $(r(r+1)/2) \times (r(r+1)/2)$ variance $\boldsymbol{V}_{\epsilon,t}$ is specified using a discount factor δ_ϵ by

$$\boldsymbol{V}_{\epsilon,t} = \frac{1-\delta_\epsilon}{\delta_\epsilon} \boldsymbol{E}_t \boldsymbol{Q}_t \boldsymbol{E}_t'.$$

Note that

$$\mathrm{E}[\boldsymbol{e}_t | D_{t-1}] = \boldsymbol{0},$$

$$\mathrm{E}[\mathrm{vech}(\boldsymbol{e}_t \boldsymbol{e}_t') | D_{t-1}] = \mathrm{vech}(\boldsymbol{Q}_t),$$

$$\mathrm{V}[\mathrm{vech}(\boldsymbol{e}_t \boldsymbol{e}_t') | D_{t-1}] = \boldsymbol{E}_t \boldsymbol{Q}_t \boldsymbol{E}_t' / \delta_\epsilon,$$

which justify the choice of the model. Of course care must be taken in specifying $\boldsymbol{E}_t$.

Theorem 5.2. *If assumptions (5.6), (5.7) are applied in the multivariate DLM (5.5), (5.5′) together with equation (5.8), then the following apply*

(a) If S_t is the mean of Σ given D_t, then

$$n_t S_t = n_{t-1} S_{t-1} + S_{t-1}^{1/2} Q_t^{-1/2} e_t e_t' Q_t^{-1/2} S_{t-1}^{1/2}, \tag{5.9}$$

$$n_t = n_{t-1} + 1,$$

(b) If V_t is the variance of vech(Σ) given D_t, then

$$V_t = V_{t-1} - (H_r(\mathcal{A}_t \otimes \mathcal{A}_t)G_r)E_t Q_t E_t'(H_r(\mathcal{A}_t \otimes \mathcal{A}_t)G_r)'/\delta_\epsilon,$$

where G_r is the duplication matrix, H_r any left inverse of it, and δ_ϵ a discount factor.

Proof. The weak assumption (5.6) implies

(i) $\mathrm{E}[\mathrm{vech}(\Sigma - \mathcal{A}_t e_t e_t' \mathcal{A}_t')|D_t] = \mathrm{E}[\mathrm{vech}(\Sigma - \mathcal{A}_t e_t e_t' \mathcal{A}_t')|D_{t-1}]$;

(ii) $\mathrm{V}[\mathrm{vech}(\Sigma - \mathcal{A}_t e_t e_t' \mathcal{A}_t')|D_t] = \mathrm{V}[\mathrm{vech}(\Sigma - \mathcal{A}_t e_t e_t' \mathcal{A}_t')|D_{t-1}]$;

(iii) $\mathrm{C}[\mathrm{vech}(\Sigma - \mathcal{A}_t e_t e_t' \mathcal{A}_t'), \mathrm{vech}(Y_t)|D_{t-1}] = \mathbf{O}$.

First we prove (a). From equation (5.8) we have

$$\mathrm{E}[e_t|D_{t-1}] = \mathbf{0},$$

$$\mathrm{E}[e_t e_t'|D_{t-1}] = \mathrm{V}[e_t|D_{t-1}] = Q_t,$$

where $Q_t = F_t' R_t F_t + S_{t-1}$. Now from (i) and $\mathrm{E}[\Sigma|D_t] = S_t$ we have

$$\begin{aligned} \mathrm{vech}(S_t) &= \mathrm{vech}(S_{t-1}) + \mathrm{vech}(\mathcal{A}_t(e_t e_t' - \mathrm{E}[e_t e_t'|D_{t-1}])\mathcal{A}_t') \\ &= \mathrm{vech}(S_{t-1} + \mathcal{A}_t(e_t e_t' - Q_t)\mathcal{A}_t'). \end{aligned}$$

Substituting $\mathcal{A}_t$ in the last equation as defined in the theorem we derive equations (5.9) as required.

From (ii) and assumption (5.7) we have

$$\mathrm{V}[\mathrm{vech}(\Sigma)|D_t] = \mathrm{V}[\mathrm{vech}(\Sigma)|D_{t-1}] - \mathrm{V}[\mathrm{vech}(\mathcal{A}_t e_t e_t' \mathcal{A}_t')|D_{t-1}]$$

and using

$$\mathrm{vech}(\mathcal{A}_t e_t e_t' \mathcal{A}_t') = \boldsymbol{H}_r(\mathcal{A}_t \otimes \mathcal{A}_t)\boldsymbol{G}_r \mathrm{vech}(e_t e_t'),$$

where $\boldsymbol{G}_r$ is the duplication matrix and $\boldsymbol{H}_r$ an arbitrary left inverse of it (see Appendix A.2), we get

$$\begin{aligned} \boldsymbol{V}_t &= \boldsymbol{V}_{t-1} - \mathrm{V}[\boldsymbol{H}_r(\mathcal{A}_t \otimes \mathcal{A}_t)\boldsymbol{G}_r \mathrm{vech}(e_t e_t')|D_{t-1}] \\ &= \boldsymbol{V}_{t-1} - (\boldsymbol{H}_r(\mathcal{A}_t \otimes \mathcal{A}_t)\boldsymbol{G}_r)\mathrm{V}[\mathrm{vech}(e_t e_t')|D_{t-1}] \\ &\quad \times (\boldsymbol{H}_r(\mathcal{A}_t \otimes \mathcal{A}_t)\boldsymbol{G}_r)' \end{aligned} \tag{5.10}$$

which with equation (5.8) provides the required result. □

Note that equation (5.10) shows that $\boldsymbol{V}_t$ is decreasing (see Appendix A.5). And since $\boldsymbol{V}_t$ is bounded below by $\mathbf{O}$ and above by $\boldsymbol{V}_0$, its limit exists. From equation (5.10) we have

$$\boldsymbol{V}_t = \boldsymbol{V}_0 - \sum_{j=1}^{t}(\boldsymbol{H}_r(\mathcal{A}_j \otimes \mathcal{A}_j)\boldsymbol{G}_r)\boldsymbol{E}_j\boldsymbol{Q}_j\boldsymbol{E}_j'(\boldsymbol{H}_r(\mathcal{A}_j \otimes \mathcal{A}_j)\boldsymbol{G}_r)'/\delta_\epsilon.$$

The existence of $\lim_{t\to\infty}\boldsymbol{V}_t$ proves that the series

$$\sum_{t=1}^{\infty}(\boldsymbol{H}_r(\mathcal{A}_t \otimes \mathcal{A}_t)\boldsymbol{G}_r)\boldsymbol{E}_t\boldsymbol{Q}_t\boldsymbol{E}_t'(\boldsymbol{H}_r(\mathcal{A}_t \otimes \mathcal{A}_t)\boldsymbol{G}_r)'/\delta_\epsilon$$

is convergent. If it is set

$$\boldsymbol{V}_0 = \sum_{t=1}^{\infty}(\boldsymbol{H}_r(\mathcal{A}_t \otimes \mathcal{A}_t)\boldsymbol{G}_r)\boldsymbol{E}_t\boldsymbol{Q}_t\boldsymbol{E}_t'(\boldsymbol{H}_r(\mathcal{A}_t \otimes \mathcal{A}_t)\boldsymbol{G}_r)'/\delta_\epsilon,$$

then

$$\lim_{t\to\infty}\boldsymbol{V}_t = \mathbf{O}.$$

5.4 Observational Variances

So far we have produced an estimate of Σ at time t, $\boldsymbol{S}_t$, such that $\mathrm{E}[\Sigma|D_t] = \boldsymbol{S}_t$ and $\mathrm{V}[\mathrm{vech}(\Sigma)|D_t] = \boldsymbol{V}_t$. So

$$(\mathrm{vech}(\Sigma)|D_t) \sim [\mathrm{vech}(\boldsymbol{S}_t), \boldsymbol{V}_t].$$

The model is closed to external information (all the information comes from the data itself; no intervention or monitoring procedures apply) and the initial distributions are partially specified as

$$(\boldsymbol{\theta}_0|D_0) \sim [\boldsymbol{m}_0, \boldsymbol{C}_0],$$
$$(\mathrm{vech}(\Sigma)|D_0) \sim [\mathrm{vech}(\boldsymbol{S}_0), \boldsymbol{V}_0],$$

for some known quantities $\boldsymbol{m}_0$, $\boldsymbol{C}_0$, $\boldsymbol{S}_0$, and $\boldsymbol{V}_0$. Employing WPP(2) modelling leads to the following assumption

$$\boldsymbol{\theta}_t - \boldsymbol{A}_t\boldsymbol{Y}_t \perp_2 \boldsymbol{Y}_t|\Sigma, D_{t-1}. \tag{5.11}$$

Theorem 5.3. *In the multivariate DLM (5.5), (5.5′) using assumption (5.11) together with the assumptions of Theorem 5.2, the one-step forecast and posterior distributions are partially derived, for each t, as follows:*

(a) Posterior at $t-1$:

For some known quantities $\boldsymbol{m}_{t-1}$, $\boldsymbol{C}_{t-1}$, $\boldsymbol{S}_{t-1}$, and $\boldsymbol{V}_{t-1}$

$$(\boldsymbol{\theta}_{t-1}|D_{t-1}) \sim [\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}],$$
$$(vech(\Sigma)|D_{t-1}) \sim [vech(\boldsymbol{S}_{t-1}), \boldsymbol{V}_{t-1}].$$

(b) Prior at t:

$$(\boldsymbol{\theta}_t|D_{t-1}) \sim [\boldsymbol{a}_t, \boldsymbol{R}_t],$$

where

$$a_t = G_t m_{t-1} \qquad \textit{and} \qquad R_t = G_t C_{t-1} G_t' + W_t.$$

(c) One-step forecast:

$$(Y_t | D_{t-1}) \sim [f_t, Q_t],$$

where

$$f_t = F_t' a_t \qquad \textit{and} \qquad Q_t = F_t' R_t F_t + S_{t-1}.$$

(d) Posterior at t:

$$(\theta_t | D_t) \sim [m_t, C_t],$$
$$(vech(\Sigma) | D_t) \sim [vech(S_t), V_t],$$

with

$$m_t = a_t + A_t e_t, \qquad C_t = R_t - A_t Q_t A_t',$$
$$n_t S_t = n_{t-1} S_{t-1} + S_{t-1}^{1/2} Q_t^{-1/2} e_t e_t' Q_t^{-1/2} S_{t-1}^{1/2}, \qquad \textit{and}$$
$$V_t = V_{t-1} - (H_r(\mathcal{A}_t \otimes \mathcal{A}_t) G_r) E_t Q_t E_t' (H_r(\mathcal{A}_t \otimes \mathcal{A}_t) G_r)' / \delta_\epsilon,$$

where

$$A_t = R_t F_t Q_t^{-1}, \qquad e_t = Y_t - f_t, \qquad n_t = n_{t-1} + 1,$$

and the quantities G_r, H_r, $\mathcal{A}_t$, E_t, δ_ϵ are as defined in Theorem 5.2.

Proof. The proof is by induction. Suppose that (a) is true. (b) follows directly from the system equation $\theta_t = G_t \theta_{t-1} + \omega_t$, $\omega_t \sim [0, W_t]$, and the posterior $(\theta_{t-1} | D_{t-1}) \sim [m_{t-1}, C_{t-1}]$. (c) follows from the observation equation $Y_t = F_t' \theta_t + \nu_t$, $\nu_t \sim [0, \Sigma]$ noting that $\mathrm{V}[\nu_t | D_{t-1}] = S_{t-1}$. From the observation equation and (b) we have $\mathrm{C}[\theta_t, Y_t | D_{t-1}] = R_t F_t = A_t Q_t$, which gives the updating for A_t of (d). Now to prove the remaining of (d) we use the model statement (5.11), that is

(i) $\mathrm{E}[\theta_t - A_t Y_t|\Sigma, D_t] = \mathrm{E}[\theta_t - A_t Y_t|\Sigma, D_{t-1}]$;

(ii) $\mathrm{V}[\theta_t - A_t Y_t|\Sigma, D_t] = \mathrm{V}[\theta_t - A_t Y_t|\Sigma, D_{t-1}]$;

(iii) $\mathrm{C}[\theta_t - A_t Y_t, Y_t|\Sigma, D_{t-1}] = \mathbf{O}$.

From (i) we have

$$\mathrm{E}[\theta_t|\Sigma, D_t] = \mathrm{E}[\theta_t|\Sigma, D_{t-1}] + A_t(Y_t - \mathrm{E}[Y_t|\Sigma, D_{t-1}]) \tag{5.12}$$

and using the first equation of Theorem 5.1 $m_t = a_t + A_t e_t$. Equation (5.12) leads to

$$\mathrm{V}[\mathrm{E}(\theta_t|\Sigma, D_t)] = (I - A_t F_t')\mathrm{V}[\mathrm{E}(\theta_t|\Sigma, D_{t-1})](I - F_t A_t'). \tag{5.13}$$

Similarly from (ii)

$$\begin{aligned}
\mathrm{V}[\theta_t|\Sigma, D_t] &= \mathrm{V}[\theta_t|\Sigma, D_{t-1}] + A_t F_t' \mathrm{V}[\theta_t|\Sigma, D_{t-1}] F_t A_t' \\
&\quad -2\mathrm{V}[\theta_t|\Sigma, D_{t-1}] F_t A_t' + A_t \mathrm{V}[\nu_t|\Sigma, D_{t-1}] A_t' \\
&= (I - A_t F_t')\mathrm{V}[\theta_t|\Sigma, D_{t-1}](I - F_t A_t') + A_t \Sigma A_t'
\end{aligned}$$

and taking the mean of the last expression

$$\mathrm{E}[\mathrm{V}(\theta_t|\Sigma, D_t)] = (I - A_t F_t')\mathrm{E}[\mathrm{V}(\theta_t|\Sigma, D_{t-1})](I - F_t A_t') + A_t S_{t-1} A_t'. \tag{5.14}$$

Equations (5.13), (5.14) together with the second one of Theorem 5.1 give

$$\begin{aligned}
C_t &= (I - A_t F_t') R_t (I - F_t A_t') + A_t S_{t-1} A_t' \\
&= R_t - A_t Q_t A_t'.
\end{aligned}$$

The updating of the moments of $(\mathrm{vech}(\Sigma)|D_t)$ is the result of Theorem 5.2. The theorem is true for $t = 1$ and so by induction is true for all $t \geq 1$. $\square$

Theorem 5.4. *For each time t and $k \geq 1$, under the conditions of Theorem 5.3, the k-step ahead distributions for $\boldsymbol{\theta}_{t+k}$, and $\boldsymbol{Y}_{t+k}$, given D_t are partially derived by*

(a) State distribution: $(\boldsymbol{\theta}_{t+k}|D_t) \sim [\boldsymbol{a}_t(k), \boldsymbol{R}_t(k)]$,

(b) Forecast distribution: $(\boldsymbol{Y}_{t+k}|D_t) \sim [\boldsymbol{f}_t(k), \boldsymbol{Q}_t(k)]$,

with moments defined recursively by

$$\boldsymbol{f}_t(k) = \boldsymbol{F}'_{t+k}\boldsymbol{a}_t(k) \qquad \textit{and} \qquad \boldsymbol{Q}_t(k) = \boldsymbol{F}'_{t+k}\boldsymbol{R}_t(k)\boldsymbol{F}_{t+k} + \boldsymbol{S}_t,$$

where

$$\boldsymbol{a}_t(k) = \boldsymbol{G}_{t+k}\boldsymbol{a}_t(k-1) \qquad \textit{and} \qquad \boldsymbol{R}_t(k) = \boldsymbol{G}_{t+k}\boldsymbol{R}_t(k-1)\boldsymbol{G}'_{t+k} + \boldsymbol{W}_{t+k},$$

with starting values

$$\boldsymbol{a}_t(0) = \boldsymbol{m}_t \qquad \textit{and} \qquad \boldsymbol{R}_t(0) = \boldsymbol{C}_t.$$

Proof. Define the $n \times n$ matrices $\boldsymbol{H}_{t+k}(x) = \boldsymbol{G}_{t+k}\boldsymbol{G}_{t+k-1}\dots\boldsymbol{G}_{t+k-x+1}$, for all t and integer $0 \leq x \leq k$, with $\boldsymbol{H}_{t+k}(0) = \boldsymbol{I}$. It is $\boldsymbol{H}_{t+k}(x) = \boldsymbol{G}_{t+k}\boldsymbol{H}_{t+k-1}(x-1)$, $1 \leq x \leq k$. From repeated application of the state evolution equation,

$$\boldsymbol{\theta}_{t+k} = \boldsymbol{H}_{t+k}(k)\boldsymbol{\theta}_t + \sum_{x=1}^{k}\boldsymbol{H}_{t+k}(k-x)\boldsymbol{\omega}_{t+x}.$$

Thus from Theorem 5.3

$$\boldsymbol{a}_t(k) = \mathrm{E}[\boldsymbol{\theta}_{t+k}|D_t] = \boldsymbol{H}_{t+k}(k)\boldsymbol{m}_t = \boldsymbol{G}_{t+k}\boldsymbol{a}_t(k-1),$$

and

$$\begin{aligned}\boldsymbol{R}_t(k) = \mathrm{V}[\boldsymbol{\theta}_{t+k}|D_t] &= \boldsymbol{H}_{t+k}(k)\boldsymbol{C}_t\boldsymbol{H}'_{t+k}(k) \\ &\quad + \sum_{x=1}^{k}\boldsymbol{H}_{t+k}(k-x)\boldsymbol{W}_{t+x}\boldsymbol{H}'_{t+k}(k-x) \\ &= \boldsymbol{G}_{t+k}\boldsymbol{R}_t(k-1)\boldsymbol{G}'_{t+k} + \boldsymbol{W}_{t+k},\end{aligned}$$

with starting values $\boldsymbol{a}_t(0) = \boldsymbol{m}_t$, and $\boldsymbol{R}_t(0) = \boldsymbol{C}_t$. This establishes (a).

Using now the observation equation at time $t+k$ and (a), $\mathrm{E}[\boldsymbol{Y}_{t+k}|D_t] = \boldsymbol{F}'_{t+k}\boldsymbol{a}_t(k)$, and $\mathrm{V}[\boldsymbol{Y}_{t+k}|D_t] = \boldsymbol{F}'_{t+k}\boldsymbol{R}_t(k)\boldsymbol{F}_{t+k} + \boldsymbol{S}_t$, as required. $\square$

The next theorem provides the retrospective or filtered distributions backwards in time. For $k \geq 1$ the definition of the k-step ahead state forecast distributions with moments $\boldsymbol{a}_t(k)$ and $\boldsymbol{R}_t(k)$ is extended to negative arguments $\boldsymbol{a}_t(-k)$ and $\boldsymbol{R}_t(-k)$.

Theorem 5.5. *In the multivariate DLM (5.5), (5.5′), for all t, define*

$$\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{G}'_{t+1}\boldsymbol{R}^{-1}_{t+1}.$$

If assumptions of Theorem 5.3 together with

$$\boldsymbol{\theta}_{t-k} - \boldsymbol{B}_{t-k}\boldsymbol{\theta}_{t-k+1} \perp_1 \boldsymbol{\theta}_{t-k+1}|D_{t-k}, \tag{5.15}$$

are applied, then for all k, $(1 \leq k \leq t-1)$, the filtered marginal distributions are partially derived by

$$(\boldsymbol{\theta}_{t-k}|D_t) \sim [\boldsymbol{a}_t(-k), \boldsymbol{R}_t(-k)],$$

where

$$\boldsymbol{a}_t(-k) = \boldsymbol{m}_{t-k} + \boldsymbol{B}_{t-k}[\boldsymbol{a}_t(-k+1) - \boldsymbol{a}_{t-k+1}],$$

and

$$\boldsymbol{R}_t(-k) = \boldsymbol{C}_{t-k} + \boldsymbol{B}_{t-k}[\boldsymbol{R}_t(-k+1) - \boldsymbol{R}_{t-k+1}]\boldsymbol{B}'_{t-k},$$

with starting values

$$\boldsymbol{a}_t(0) = \boldsymbol{m}_t \qquad \textit{and} \qquad \boldsymbol{R}_t(0) = \boldsymbol{C}_t,$$

and

$$\boldsymbol{a}_{t-k}(1) = \boldsymbol{a}_{t-k+1} \qquad \textit{and} \qquad \boldsymbol{R}_{t-k}(1) = \boldsymbol{R}_{t-k+1}.$$

Proof. The filtered densities are defined by

$$p(\boldsymbol{\theta}_{t-k}|D_t) = \int p(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_t)p(\boldsymbol{\theta}_{t-k+1}|D_t)\, d\boldsymbol{\theta}_{t-k+1}. \tag{5.16}$$

Assume the theorem true for $k-1$, so that

$$(\boldsymbol{\theta}_{t-k+1}|D_t) \sim [\boldsymbol{a}_t(-k+1), \boldsymbol{R}_t(-k+1)].$$

Using Bayes' theorem, the first integrand term of (5.16) is

$$p(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_t] = \frac{p(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_{t-k})p(\boldsymbol{Y}|\boldsymbol{\theta}_{t-k}, \boldsymbol{\theta}_{t-k+1}, D_{t-k})}{p(\boldsymbol{Y}|\boldsymbol{\theta}_{t-k+1}, D_{t-k})},$$

where $\boldsymbol{Y} = \{\boldsymbol{Y}_{t-k+1}, \ldots, \boldsymbol{Y}_t\}$. Now given $\boldsymbol{\theta}_{t-k+1}$, $\boldsymbol{Y}$ is independent of the previous value $\boldsymbol{\theta}_{t-k}$ so that the two terms $p(\boldsymbol{Y}|\cdot)$ cancel. Thus

$$p(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_t) = p(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_{t-k}). \tag{5.17}$$

By Bayes' theorem the right hand side term is

$$p(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_{t-k}) \propto p(\boldsymbol{\theta}_{t-k}|D_{t-k})p(\boldsymbol{\theta}_{t-k+1}|\boldsymbol{\theta}_{t-k}, D_{t-k}). \tag{5.18}$$

Now

$$(\boldsymbol{\theta}_{t-k+1}|\boldsymbol{\theta}_{t-k}, D_{t-k}) \sim [\boldsymbol{G}_{t-k+1}\boldsymbol{\theta}_{t-k}, \boldsymbol{W}_{t-k+1}],$$

and from Theorem 5.3

$$(\boldsymbol{\theta}_{t-k}|D_{t-k}) \sim [\boldsymbol{m}_{t-k}, \boldsymbol{C}_{t-k}].$$

Form the joint distribution $p(\boldsymbol{\theta}_{t-k}, \boldsymbol{\theta}_{t-k+1}|D_{t-k})$ as

$$\left(\begin{array}{c} \boldsymbol{\theta}_{t-k} \\ \boldsymbol{\theta}_{t-k+1} \end{array} \middle| D_{t-k} \right) \sim \left[\begin{pmatrix} \boldsymbol{m}_{t-k} \\ \boldsymbol{a}_{t-k+1} \end{pmatrix}, \begin{pmatrix} \boldsymbol{C}_{t-k} & \boldsymbol{B}_{t-k}\boldsymbol{R}_{t-k+1} \\ \boldsymbol{R}_{t-k+1}\boldsymbol{B}'_{t-k} & \boldsymbol{R}_{t-k+1} \end{pmatrix} \right].$$

Note that $\mathrm{C}[\boldsymbol{\theta}_{t-k}, \boldsymbol{\theta}_{t-k+1} | D_{t-k}] = \boldsymbol{C}_{t-k}\boldsymbol{G}'_{t-k+1} = \boldsymbol{B}_{t-k}\boldsymbol{R}_{t-k+1}$, where $\boldsymbol{B}_{t-k}$ is the regression matrix of $\boldsymbol{\theta}_{t-k}$ on $\boldsymbol{\theta}_{t-k+1}$. From assumption (5.15) and Theorem 5.1 it is easily deduced that

$$\mathrm{E}[\boldsymbol{\theta}_{t-k} | \boldsymbol{\theta}_{t-k+1}, D_{t-k}] = \boldsymbol{m}_{t-k} + \boldsymbol{B}_{t-k}(\boldsymbol{\theta}_{t-k+1} - \boldsymbol{a}_{t-k+1}) = \boldsymbol{l}_t(k),$$

and

$$\mathrm{V}[\boldsymbol{\theta}_{t-k} | \boldsymbol{\theta}_{t-k+1}, D_{t-k}] = \boldsymbol{C}_{t-k} - \boldsymbol{B}_{t-k}\boldsymbol{R}_{t-k+1}\boldsymbol{B}'_{t-k} = \boldsymbol{L}_t(k).$$

Applying again Theorem 5.1 in (5.17), the moments of $(\boldsymbol{\theta}_{t-k} | D_t)$ are

$$\begin{aligned} \mathrm{E}[\boldsymbol{\theta}_{t-k} | D_t] &= \mathrm{E}[\boldsymbol{l}_t(k) | D_t] = \boldsymbol{a}_t(-k) \qquad \text{and} \\ \mathrm{V}[\boldsymbol{\theta}_{t-k} | D_t] &= \mathrm{E}[\boldsymbol{L}_t(k) | D_t] + \mathrm{V}[\boldsymbol{l}_t(k) | D_t] = \boldsymbol{R}_t(-k), \end{aligned}$$

with $\boldsymbol{a}_t(-k)$, $\boldsymbol{R}_t(-k)$ as stated in the theorem. The theorem is completed by induction, since it is true for $k = 1$ with

$$\boldsymbol{a}_t(-k+1) = \boldsymbol{a}_t(0) = \boldsymbol{m}_t$$

and

$$\boldsymbol{R}_t(-k+1) = \boldsymbol{R}_t(0) = \boldsymbol{C}_t.$$

□

Corollary 5.1. *The corresponding smoothed distribution for the mean response of the series are partially given by*

$$(\mu_{t-k} | D_t) \sim [\boldsymbol{f}_t(-k), \boldsymbol{F}'_{t-k}\boldsymbol{R}_t(-k)\boldsymbol{F}_{t-k}],$$

where $\boldsymbol{f}_t(-k) = \boldsymbol{F}'_{t-k}\boldsymbol{a}_t(-k)$.

5.5 Limiting Results for Constant Observable Models

In this section general methods are developed for the limiting behaviour of constant multivariate DLMs. Results are in parallel to [17]. There, it is mentioned that limiting results can be more general and may not rely upon normal distributions. P.J. Harrison states in this paper, "It should be noted that these results are based solely on the forms of the updating equations in TSDLMs, with absolutely no assumptions about a "true" data generating process or about normality." Similar comments can be found in [51, page 166]. However, such a development is not possible until a general DLM analysis, which provides recurrence relationships for $\boldsymbol{m}_t$ and $\boldsymbol{C}_t$, is available. Especially in the observational variances case for multivariate DLMs, such limiting results cannot be derived from the existing literature. This section uses the weak assumptions of Theorems 5.2, 5.3 so that general limiting results are derived.

Consider the multivariate constant observable DLM $\{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{\Sigma}, \boldsymbol{W}\}$

$$\boldsymbol{Y}_t = \boldsymbol{F}'\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim [0, \boldsymbol{\Sigma}], \tag{5.19}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim [0, \boldsymbol{W}], \tag{5.19$'$}$$

with components as defined in Section 5.4 and $\boldsymbol{F}_t = \boldsymbol{F}$, $\boldsymbol{G}_t = \boldsymbol{G}$, $\boldsymbol{W}_t = \boldsymbol{W}$. $\boldsymbol{\Sigma}$ is assumed unknown, but subject to the variance learning of Theorem 5.2.

Theorem 5.6. *Employing the assumptions of Theorem 5.3 and assuming that for any observable DLM additional information decreases the state vari-*

ance C_t, the limiting variance

$$\lim_{t\to\infty} C_t = C$$

exists and is independent of the initial information D_0.

Proof. The proof is identical to [17] after noticing that Theorem 5.3 holds.

□

Corollary 5.2. *Under the assumptions of Theorem 5.6, the next limiting results hold*

$$\lim_{t\to\infty} R_t = GCG' + W = R,$$

$$\lim_{t\to\infty} Q_t = F'RF + \Sigma = Q,$$

$$\lim_{t\to\infty} A_t = RFQ^{-1} = A,$$

where $\Sigma = \lim_{t\to\infty} S_t$ and S_t is the estimate of Σ at time t.

Proof. Write V_t as in Theorem 5.2. From Theorem 5.6 we can choose

$$V_0 = \sum_{t=1}^{\infty}(H_r(\mathcal{A}_t \otimes \mathcal{A}_t)G_r)E_tQ_tE_t'(H_r(\mathcal{A}_t \otimes \mathcal{A}_t)G_r)'/\delta_\epsilon,$$

with components as defined in Theorem 5.2, such that $\lim_{t\to\infty} V_t = O$. This implies that the limiting value of S_t will be the real matrix Σ. The remaining of the corollary is straight-forward from Theorems 5.3 and 5.6. □

Now using the updating of m_t of Theorem 5.3 we get

$$m_t = Gm_{t-1} + A_te_t = H_tm_{t-1} + A_tY_t,$$

where $H_t = (I - A_tF')G = (I - A_tQ_tA_t'R_t^{-1})G = C_tR_t^{-1}G$.

Assuming that the matrix $\boldsymbol{H} = \lim_{t\to\infty} \boldsymbol{H}_t$ has all its eigenvalues in the interval $(-1, 1)$ and that $\lim_{t\to\infty} \boldsymbol{Y}_t = \boldsymbol{Y}$ exists, the limit of $\boldsymbol{m}_t$ is obtained as

$$\lim_{t\to\infty} \boldsymbol{m}_t = (\boldsymbol{I} - \boldsymbol{H})^{-1}\boldsymbol{A}\boldsymbol{Y}, \tag{5.20}$$

where $\boldsymbol{H} = \boldsymbol{C}\boldsymbol{R}^{-1}\boldsymbol{G}$, see equation (A.3) in Appendix A.5.

Now from Theorem 5.5, the limit of $\boldsymbol{B}_{t-k}$ is

$$\lim_{t\to\infty} \boldsymbol{B}_{t-k} = \boldsymbol{C}\boldsymbol{G}'\boldsymbol{R}^{-1} = \boldsymbol{B}, \tag{5.21}$$

for every fixed $k > 0$.

The next theorem provides the filtered limit results for the constant DLM.

Theorem 5.7. *For any $0 < k < t$, under the conditions of Theorem 5.5, assuming that $\lim_{t\to\infty} \boldsymbol{Y}_t$ exists and the eigenvalues of $\lim_{t\to\infty} \boldsymbol{C}_t\boldsymbol{R}_t^{-1}\boldsymbol{G}$, λ_i $(i = 1, \ldots, n)$, satisfy $|\lambda_i| < 1$, the limiting filtering distribution of $(\boldsymbol{\theta}_{t-k}|D_t)$ is partially given by*

$$\lim_{t\to\infty} (\boldsymbol{\theta}_{t-k}|D_t) \sim [\boldsymbol{a}(-k), \boldsymbol{R}(-k)],$$

with moments calculated recursively by

$$\boldsymbol{a}(-k) = \boldsymbol{m} + \boldsymbol{B}[\boldsymbol{a}(-k+1) - \boldsymbol{a}],$$
$$\boldsymbol{R}(-k) = \boldsymbol{C} + \boldsymbol{B}[\boldsymbol{R}(-k+1) - \boldsymbol{R}]\boldsymbol{B}',$$

with

$$\boldsymbol{a} = \boldsymbol{G}(\boldsymbol{I} - \boldsymbol{H})^{-1}\boldsymbol{A}\boldsymbol{Y}, \quad \boldsymbol{a}(0) = \boldsymbol{m} = (\boldsymbol{I} - \boldsymbol{H})^{-1}\boldsymbol{A}\boldsymbol{Y} \quad \text{and} \qquad \boldsymbol{R}(0) = \boldsymbol{C},$$

where $\boldsymbol{C}_t \to \boldsymbol{C}$, $\boldsymbol{R}_t \to \boldsymbol{R}$, $\boldsymbol{A}_t \to \boldsymbol{A}$, $\boldsymbol{H}_t \to \boldsymbol{H} = \boldsymbol{C}\boldsymbol{R}^{-1}\boldsymbol{G}$, and $\boldsymbol{Y}_t \to \boldsymbol{Y}$, as $t \to \infty$.

Proof. The proof is straight-forward from Theorem 5.5, Corollary 5.2, and equations (5.20), (5.21). □

If $\boldsymbol{H}$ has at least one eigenvalue λ_j with $|\lambda_j| \geq 1$, or if there is evidence that $\lim_{t\to\infty} \boldsymbol{Y}_t$ does not exist (e.g. seasonal time series), then $\boldsymbol{a}(-k)$ of the above theorem can be approximated by

$$\boldsymbol{a}_t(-k) \approx \boldsymbol{m}_{t-k} + \boldsymbol{B}[\boldsymbol{a}_t(-k+1) - \boldsymbol{a}_{t-k+1}],$$

where $\boldsymbol{a}_t(0) = \boldsymbol{m}_t$ and $\boldsymbol{a}_{t-1}(1) = \boldsymbol{a}_t$.

5.6 Illustration

In this section the model developed in Sections 5.3 and 5.4 is illustrated and a comparison with existing models is attempted. For this purpose 2 bivariate series are generated.

Series a ($\boldsymbol{Y}_t^{(a)}$) and Series b ($\boldsymbol{Y}_t^{(b)}$) are shown in Table C.1 (page 238). For both series the general multivariate DLM, given by equations (5.5), (5.5′) in Section 5.3, has been used.

In generating 100 observations of $\boldsymbol{Y}_t^{(a)}$ the settings were

$$\Sigma = \begin{pmatrix} 2 & 1.5 \\ 1.5 & 4 \end{pmatrix}, \quad \boldsymbol{W}_t = \boldsymbol{W} = \begin{pmatrix} 1 & 2 \\ 2 & 6 \end{pmatrix}, \quad \boldsymbol{F}_t = \boldsymbol{F} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix},$$

$$\boldsymbol{G}_t = \boldsymbol{J}_2(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{C}_0 = \begin{pmatrix} 1000 & 0 \\ 0 & 1000 \end{pmatrix}, \quad \boldsymbol{m}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Figure 5.1 show the simulated series $\boldsymbol{Y}_t^{(a)} = (Y_{1t}^{(a)}, Y_{2t}^{(a)})'$. The continuous line corresponds to the series $Y_{1t}^{(a)}$ and the dotted line to the series $Y_{2t}^{(a)}$.

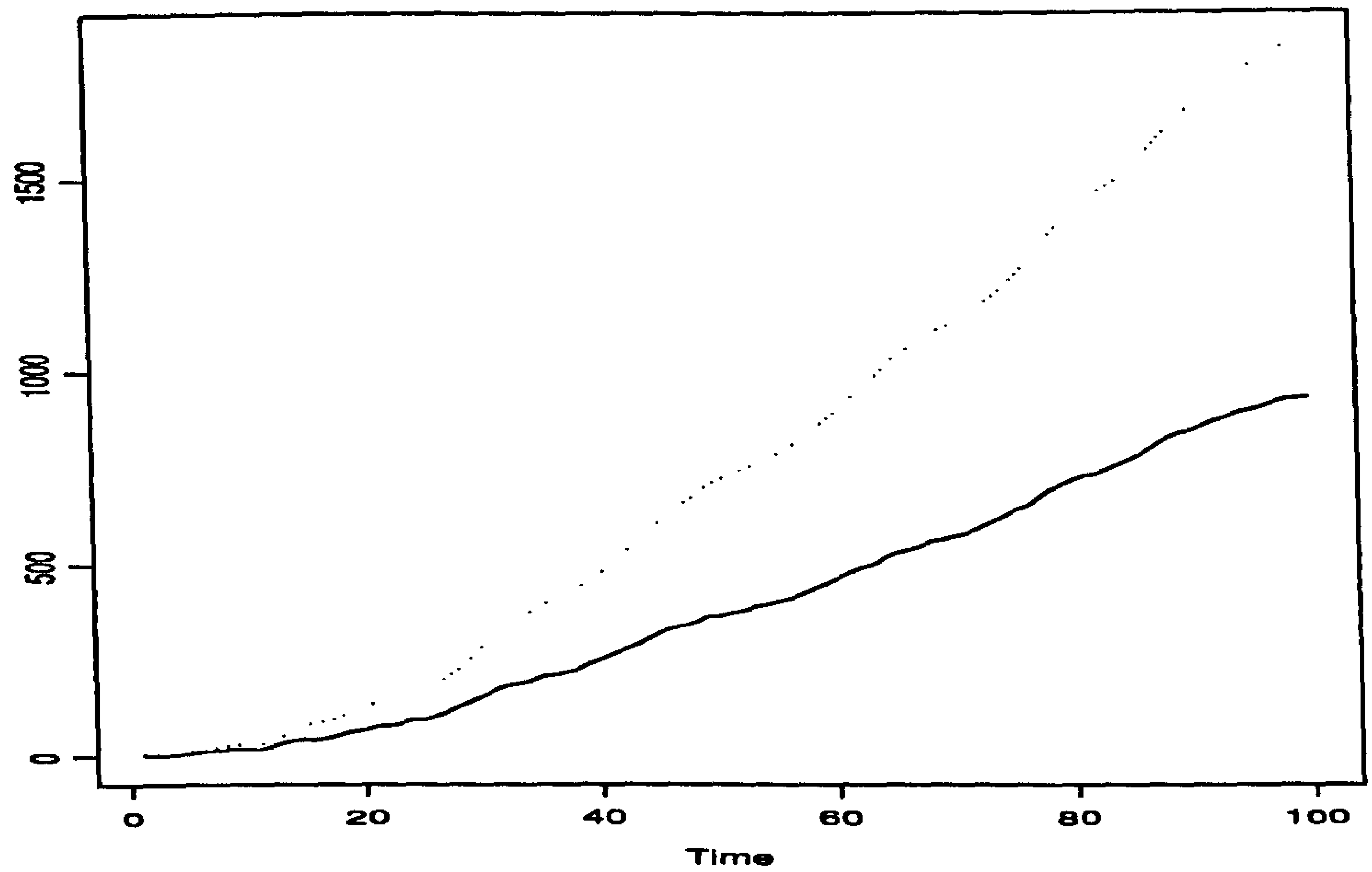

Figure 5.1: Bivariate simulated series: Series a

In order to produce the estimate of Σ, S_t, further the initial values of S_0, n_0 (n_t are the degrees of freedom at time t) were set to

$$S_0 = \begin{pmatrix} 10000 & 0 \\ 0 & 10000 \end{pmatrix}, \qquad \text{and} \qquad n_0 = \frac{1}{1000}.$$

Note that the above settings of C_0, S_0, n_0 are chosen such that no precise initial information is assumed.

The estimate of Section 5.3 (see equation (5.9)) is evaluated against the estimate of the scaled multivariate DLM approximation (see Section 4.7 and [4]). Let $\Sigma = \{\sigma_{ij}\}$ $(i, j = 1, 2)$ and $Y_t^{(a)} = (Y_{1t}^{(a)}, Y_{2t}^{(a)})'$.

To the following figures the continuous line refers to the new method (the weak probability approach of Sections 5.3 and 5.4), while the dotted line to the old method (see [4] or Section 4.7). Figures 5.2, 5.3, 5.4 show a notably better performance of the new method. At $t = 20$ the new method estimates

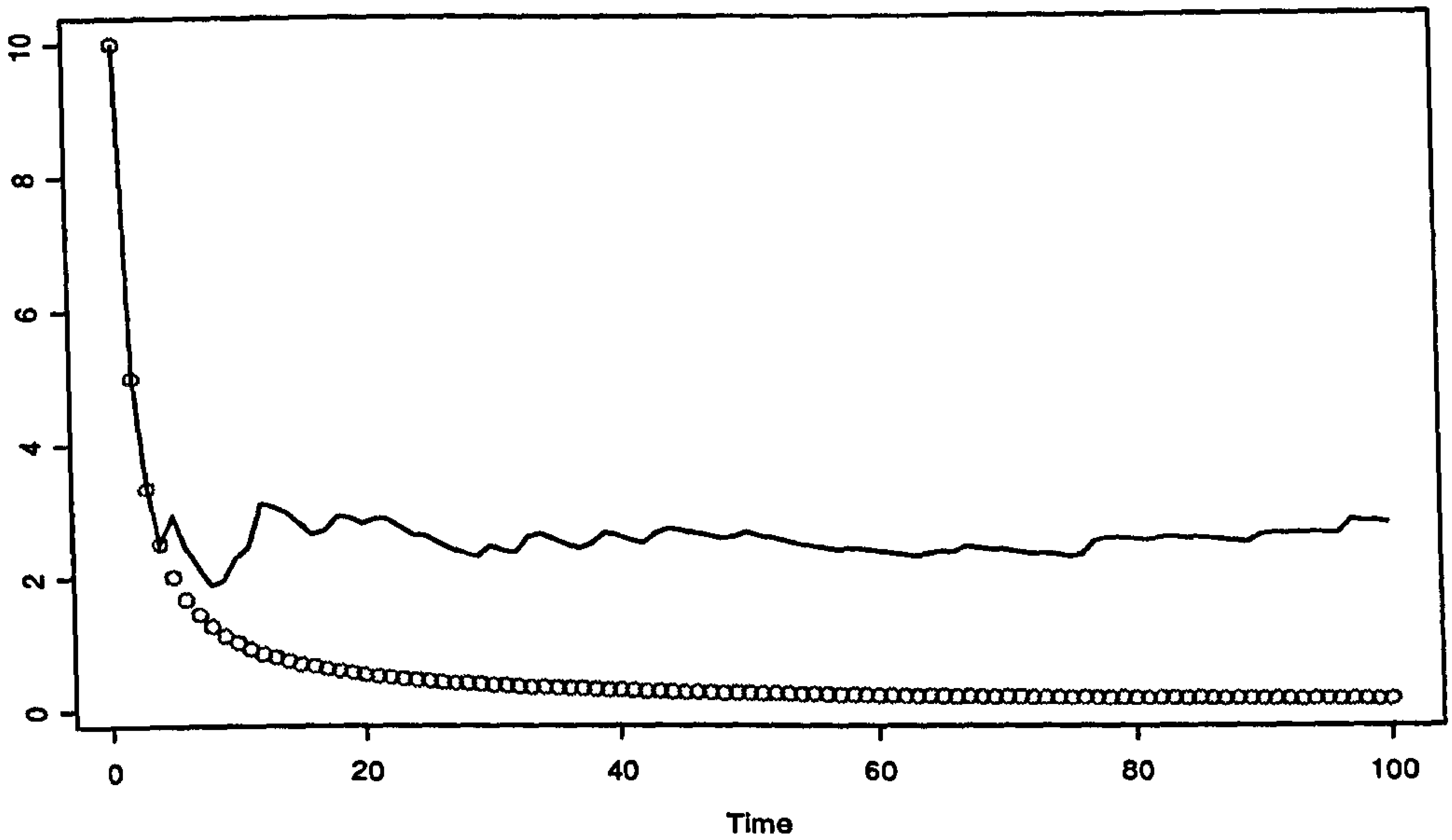

Figure 5.2: Variance estimation for $\sigma_{11} = 2$: Series a

the variance $\sigma_{11} = 2$ as 2.746 while the older method as 0.626. After this time the new method slowly improves the estimate (e.g. 2.6 at $t = 40$) while the older method is rapidly departing from it (e.g. 0.4 at $t = 40$). Figure 5.3 shows the estimate of the covariance, σ_{12}. The new method is approaching the real $\sigma_{12} = 1.5$ after $t = 22$ having there an estimate of 0.74 and is reaching the value of 0.95 at $t = 91$. The older method is around zero having several negative values and leading to a poor non-improving estimate. In fact with the last method it seems that $Y_{1t}^{(a)}$ and $Y_{2t}^{(a)}$ are uncorrelated which is not true. Similarly in Figure 5.4 we see that the new method produces much better estimate for the variance $\sigma_{22} = 4$. At $t = 20$ the new method reaches the value of 2.8, while the older method's estimate is only 0.5. Note that the older method's estimate is departing from $\sigma_{22} = 4$ as t increases.

Series $\boldsymbol{Y}_t^{(b)}$ of Table C.1 has been generated retaining all the respective

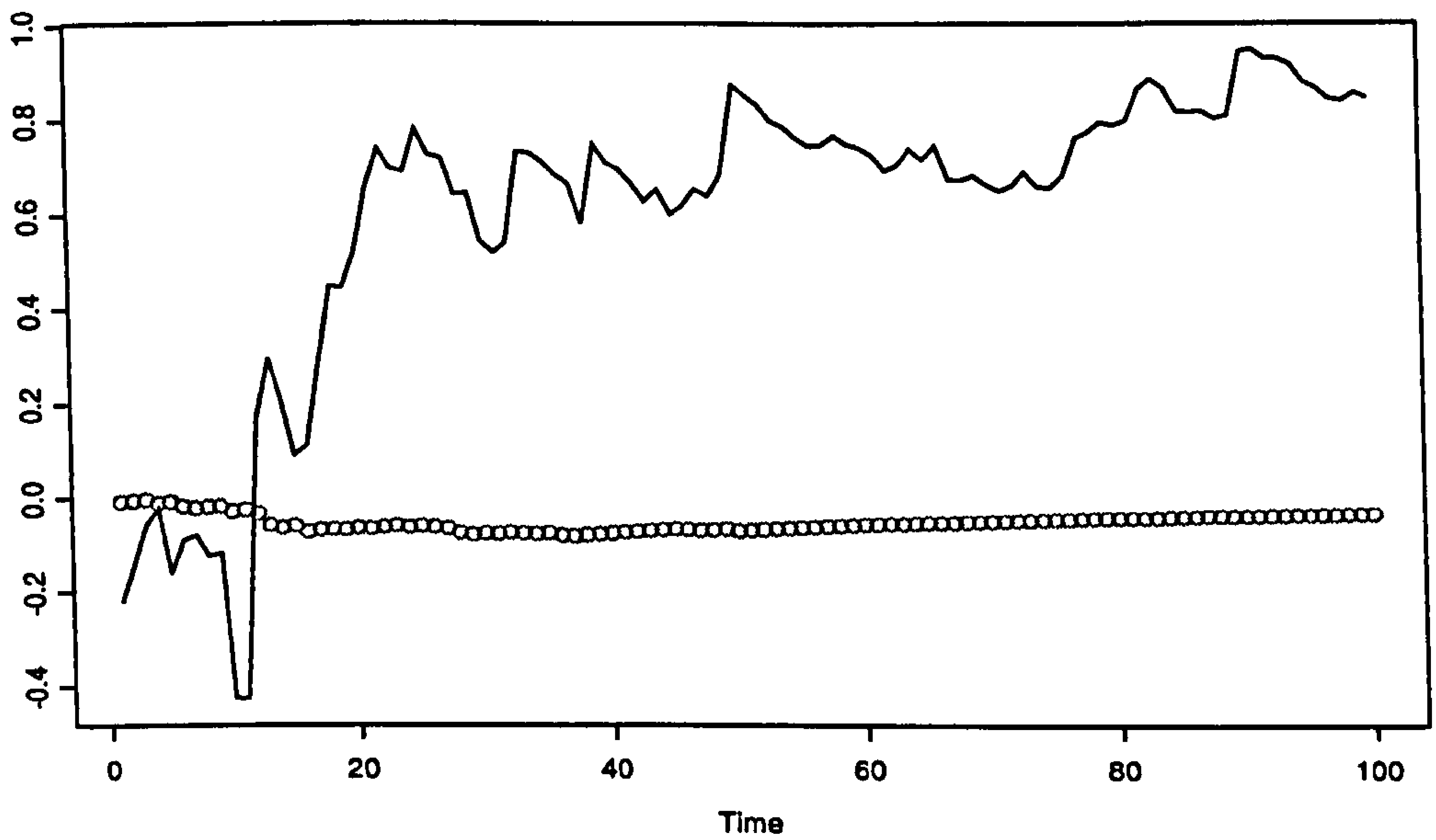

Figure 5.3: Variance estimation for $\sigma_{12} = 1.5$: Series a

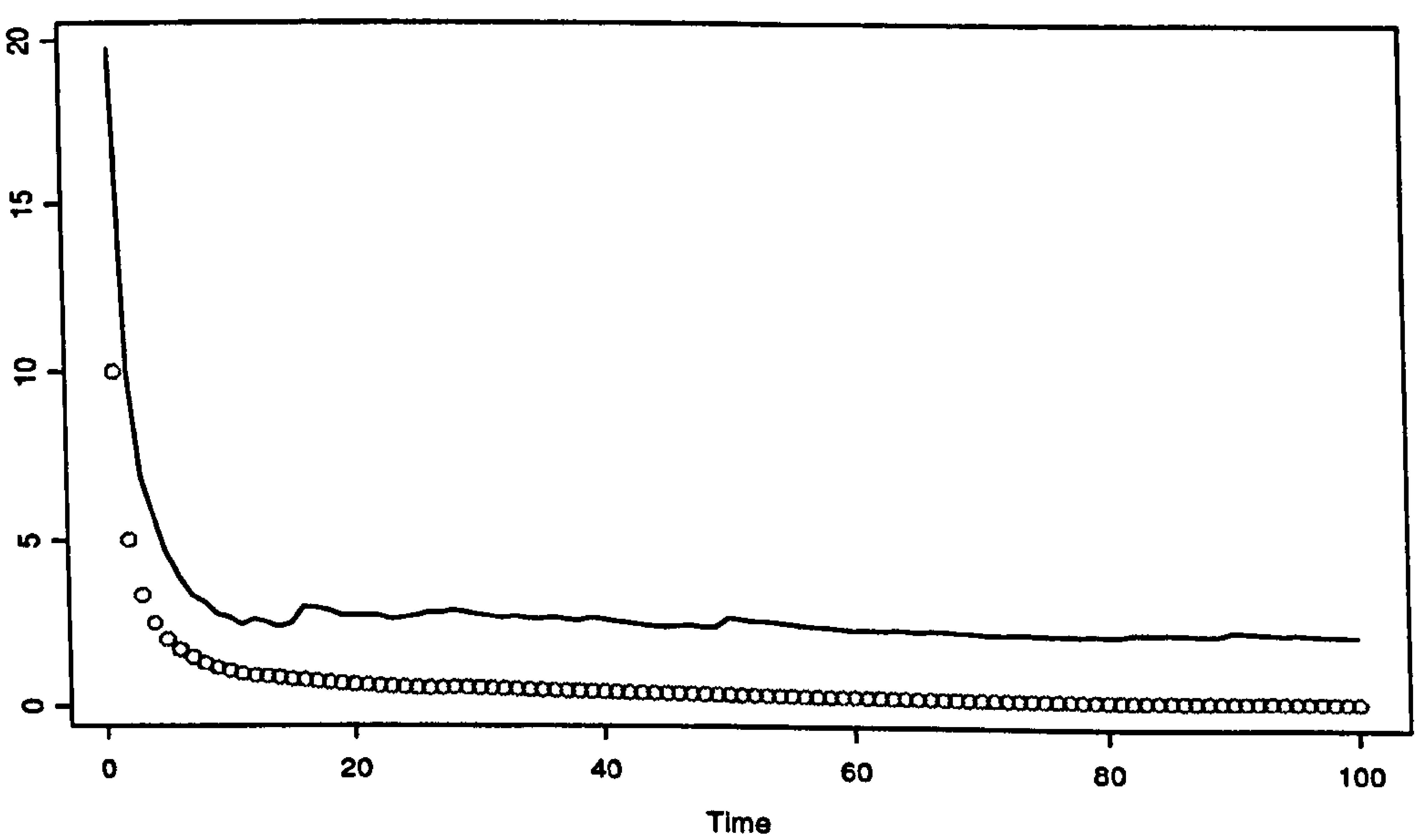

Figure 5.4: Variance estimation for $\sigma_{22} = 4$: Series a

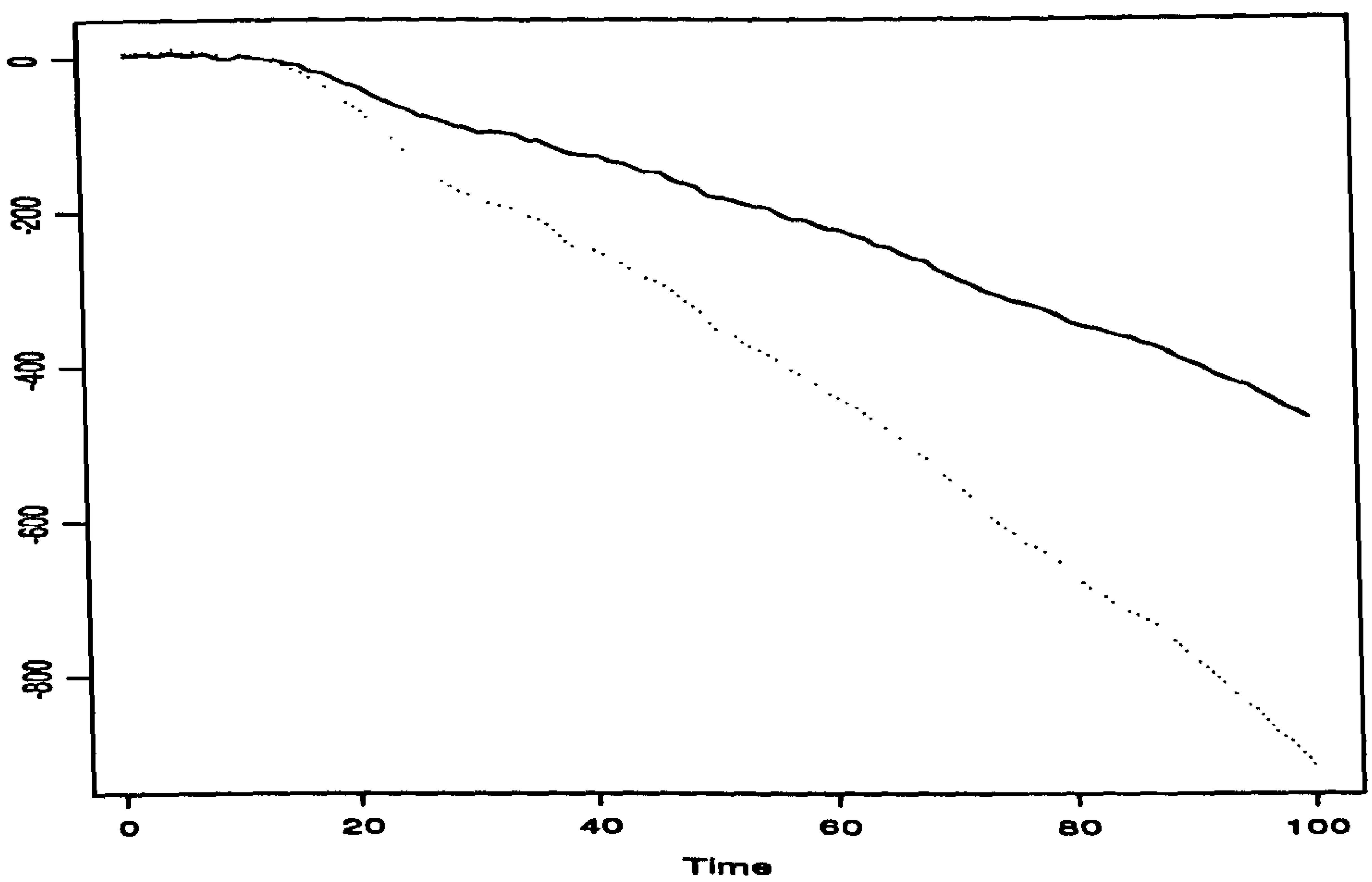

Figure 5.5: Bivariate simulated series: Series b

quantities of $Y_t^{(a)}$, but

$$W_t = W = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.7 \end{pmatrix}, \qquad C_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

By choosing a precise variance $C_0 = O$ and low values of W, we concentrate the uncertainty of the generated observations $Y_t^{(b)}$ ($t = 1, \ldots, 100$) around the observational variance matrix Σ. So we can evaluate better the performance of its estimate, S_t, produced by the two methods.

Figure 5.5 show the simulated series $Y_t^{(b)} = (Y_{1t}^{(b)}, Y_{2t}^{(b)})'$. The continuous line corresponds to the series $Y_{1t}^{(b)}$ and the dotted line to the series $Y_{2t}^{(b)}$.

The results are shown in Figures 5.6, 5.7, and 5.8. Similar comments apply. In brief, Figure 5.6 shows that the new method very soon ($t = 6$) reaches the value of 2.2, while the older method only around $t = 40$ has a good performance, which unfortunately is not retained. The older method

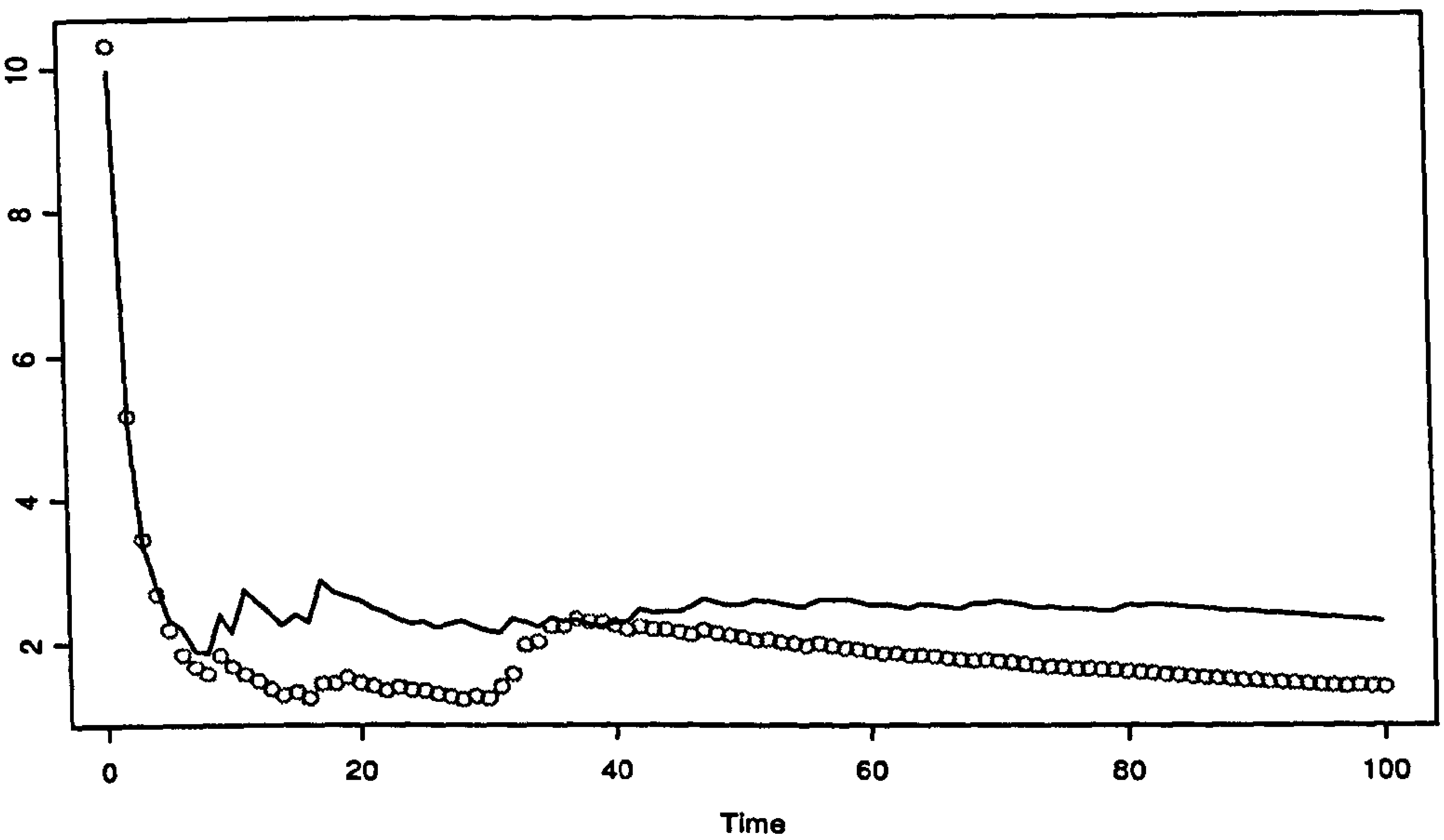

Figure 5.6: Variance estimation for $\sigma_{11} = 2$: Series b

produces a very poor covariance estimate, similarly as in Series a, around zero. Figure 5.7 shows clearly that the new method's estimate is very close to 1.5 from $t = 35$ and on. Again, Figure 5.8 shows that from $t = 10$ the new method produces an estimate for $\sigma_{22} = 4$ of 4.1 and retains it with slight jumps. However, the older method is hardly reaching the value of 2.

Similar results found on several other simulations. The new method, based on the Weak Probability approach is more general (may be applied even if normal distributions are not assumed) and the variance estimate has a better performance. The one-step forecast function and errors of the new and the old methods are identical, since no variance estimate is involved. However, the one-step variance will be different. Also note that the posterior distributions of the scaled version of the general multivariate DLM (older method) depends upon the initial variance $\boldsymbol{S}_0$. This is a question, how $\boldsymbol{S}_0$

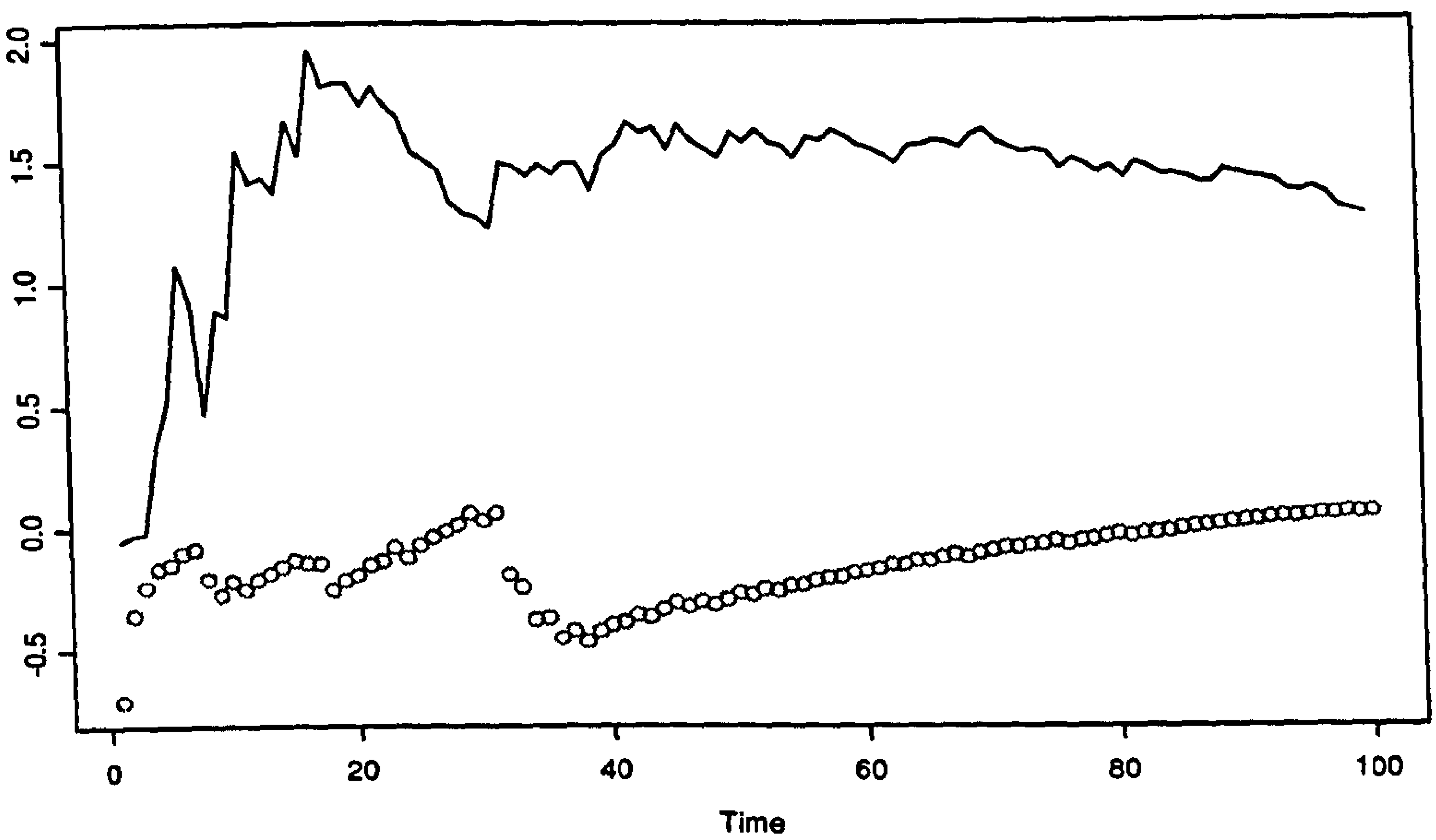

Figure 5.7: Variance estimation for $\sigma_{12} = 1.5$: Series b

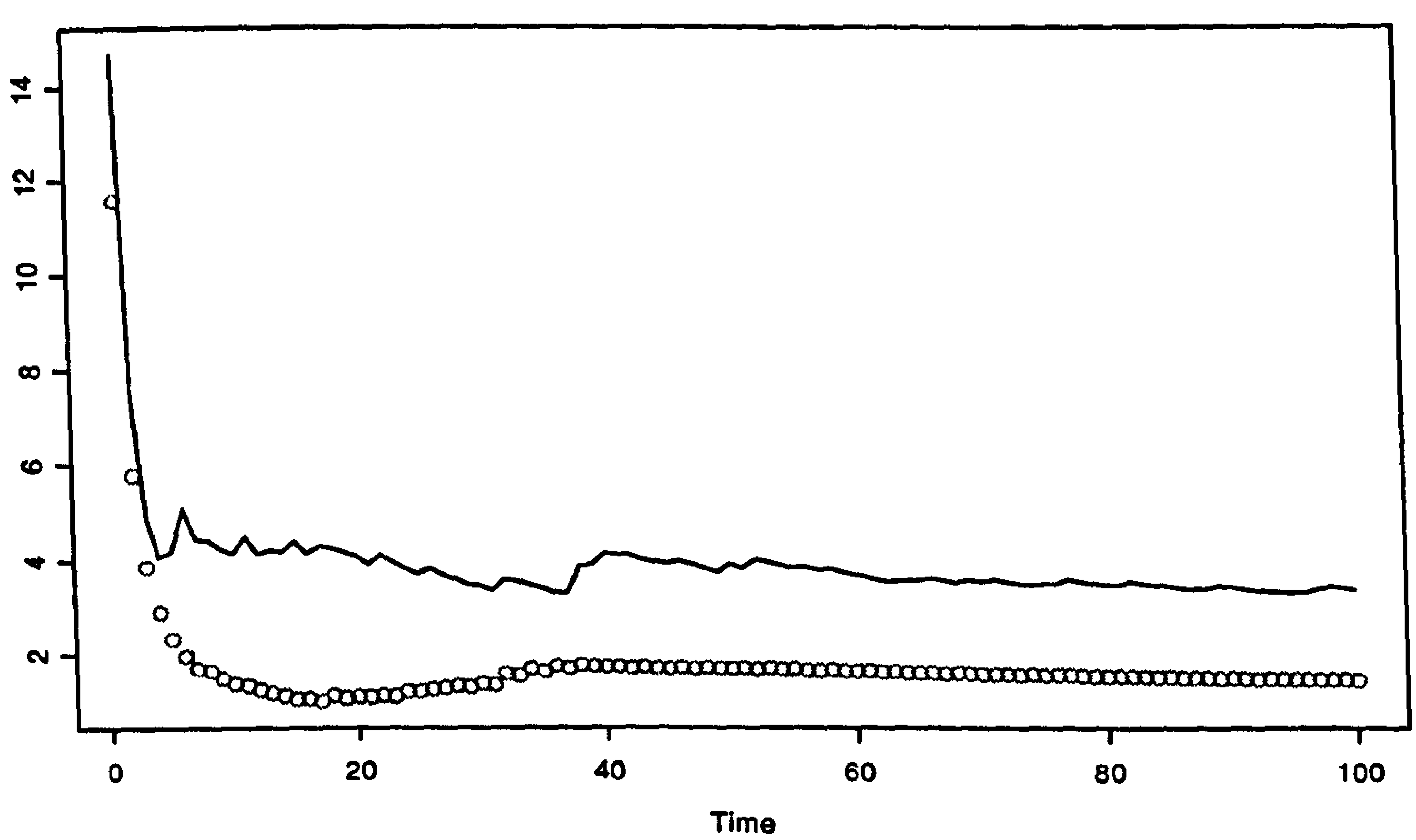

Figure 5.8: Variance estimation for $\sigma_{22} = 4$: Series b

affects the posterior distribution of the state vector. On the other hand, the method proposed in this chapter has most of the advantages of the normal DLM theory, based on known observational variances. As developed in this chapter, k-step forecasting, retrospective analysis, and limiting results, are all available within a neat context. Also, in Chapters 6, 7 some further topics of multivariate DLM theory are examined using this model. These topics include missing observations and time-varying observational variances.

All the above demonstrate that the Weak Probability approach is very useful in the arena of multivariate modelling.

CHAPTER 6

Missing Observations

6.1 Introduction

This chapter deals with missing observations. The main part of this work is to be found in [40]. Section 6.2 introduces the problem of missing observations in the multivariate models and provides a review of the existing treatment. Section 6.3 proposes a methodology for the models developed in Chapters 3, 4, and 5, including the multivariate DLM with known variances. In Section 6.4 an algorithm dealing with unequally spaced observations is presented. Finally, Section 6.5 examines an example of a time series with both unequally spaced and missing observations.

6.2 The Problem of Missing Observations

Missing observations play a significant role in almost every field in Statistics. DLMs are no exception. Missing observations arise frequently in scientific fields and also in the socio-economics.

The analysis of missing observations plays an important role in intervention. When a value is identified as an outlier and it may not be desirable that it be used for future calculations, then it should be treated as a missing observation.

In the univariate case the problem of missing observations is easily resolved. For example, assume the univariate model

$$Y_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, \qquad \nu_t \sim \mathrm{N}[0, V_t],$$
$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{W}_t],$$

where Y_t is a scalar observation variable, $\boldsymbol{F}_t$ an $n \times 1$ design vector, $\boldsymbol{\theta}_t$ an $n \times 1$ state vector, ν_t a random variable with known or unknown variance V_t, $\boldsymbol{G}_t$ a known $n \times n$ evolution matrix, and $\boldsymbol{\omega}_t$ an $n \times 1$ random vector with known evolution variance matrix $\boldsymbol{W}_t$. If at time t the observation Y_t is missing, the posterior distribution at t, $p(\boldsymbol{\theta}_t|D_t)$, is simply the prior distribution at t, $p(\boldsymbol{\theta}_t|D_{t-1})$, since no information comes in to the system at time t ($D_t = D_{t-1}$).

Such a straight forward approach in the multivariate case is restricted to the case where the whole observation vector is missing at a specific time t. If a subvector, $\boldsymbol{Y}_t^{(s)}$, of $\boldsymbol{Y}_t$ is missing the above methodology is no longer applicable. Since some values are observed, there is some form of the likelihood function. However, it is not known how the missing values affect the likelihood function.

Assuming known observationsal variances, Shumway and Stoffer, [39], propose the EM (Expectation Maximization) algorithm for multivariate state space time series. They provide an approximation of the likelihood function based only on the observed values. However, they criticise the method, since the rate of convergence is questionable and it is probably dependent on the nature of the data. The general normal multivariate DLM is

$$\boldsymbol{Y}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{V}_t], \tag{6.1}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{W}_t], \tag{6.1$'$}$$

where $\boldsymbol{Y}_t$ is an $r \times 1$ observation vector, $\boldsymbol{F}_t$ a known $n \times r$ design matrix, $\boldsymbol{\theta}_t$ an $n \times 1$ state vector, $\boldsymbol{\nu}_t$ an $r \times 1$ random vector, $\boldsymbol{V}_t$ a known $r \times r$ variance matrix, $\boldsymbol{G}_t$ an $n \times n$ evolution matrix, $\boldsymbol{\omega}_t$ an $n \times 1$ random vector, and $\boldsymbol{W}_t$ a known $n \times n$ variance matrix.

The method assumes that matrix $\boldsymbol{V}_t$ is a diagonal matrix, or that writing $\boldsymbol{\nu}_t = (\nu_{1t}, \ldots, \nu_{rt})'$ all ν_{it} are uncorrelated. To the following we show that for known $\boldsymbol{V}_t$ this is not a limitation.

Consider the following partition of equation (6.1).

$$\boldsymbol{Y}_t = \begin{pmatrix} \boldsymbol{Y}_{1t} \\ \boldsymbol{Y}_{2t} \end{pmatrix}, \quad \boldsymbol{F}_t = \begin{pmatrix} \boldsymbol{F}_{1t}, & \boldsymbol{F}_{2t} \end{pmatrix}, \quad \boldsymbol{\nu}_t = \begin{pmatrix} \boldsymbol{\nu}_{1t} \\ \boldsymbol{\nu}_{2t} \end{pmatrix},$$

$$\boldsymbol{V}_t = \begin{pmatrix} \boldsymbol{V}_{11,t} & \boldsymbol{V}_{12,t} \\ \boldsymbol{V}_{12,t}' & \boldsymbol{V}_{22,t} \end{pmatrix},$$

where $\dim(\boldsymbol{Y}_{1t}) = q \times 1$, $\dim(\boldsymbol{Y}_{2t}) = (r-q) \times 1$, $\dim(\boldsymbol{F}_{1t}) = n \times q$, $\dim(\boldsymbol{F}_{2t}) = n \times (r-q)$, $\dim(\boldsymbol{\nu}_{1t}) = q \times 1$, $\dim(\boldsymbol{\nu}_{2t}) = (r-q) \times 1$, $\dim(\boldsymbol{V}_{11,t}) = q \times q$, $\dim(\boldsymbol{V}_{12,t}) = q \times (r-q)$, and $\dim(\boldsymbol{V}_{22,t}) = (r-q) \times (r-q)$, for some $1 \leq q \leq r-1$.

Assuming that $V_{12,t} \neq \mathrm{O}$, model (6.1), (6.1′) can be written as

$$Y_t^* = F_t^{*'}\theta_t + \nu_t^*, \qquad \nu_t^* \sim \mathrm{N}[0, V_t^*], \tag{6.2}$$

$$\theta_t = G_t\theta_{t-1} + \omega_t, \qquad \omega_t \sim \mathrm{N}[0, W_t], \tag{6.2'}$$

where $Y_t^* = M_t Y_t$, $F_t^* = F_t M_t'$, $\nu_t^* = M_t \nu_t$, and

$$V_t^* = \begin{pmatrix} V_{11,t} & \mathrm{O} \\ \mathrm{O}' & V_{22,t} - V_{12,t}' V_{11,t}^{-1} V_{12,t} \end{pmatrix},$$

with

$$M_t = \begin{pmatrix} I & \mathrm{O} \\ -V_{12,t}' V_{11,t}^{-1} & I \end{pmatrix}.$$

It is a matter of some simple algebra to verify this result. Notice that M_t is non-singular, since $|M_t| = 1$ and the diagonal elements of V_t^* are not independent. This is no problem since all $V_{11,t}$, $V_{12,t}$, $V_{22,t}$ are known.

From the above it is clear that the method is not applicable when the observational variance matrix V_t is unknown.

A much simpler and straight forward approach is proposed in [10, page 92]. Consider the above partition of Y_t, F_t, V_t, and assume that V_t is known for all t. Let Y_{1t} consists of all the observed values and Y_{2t} consists of all the missing values at t. Then working with the initial model, if $V_{12} = \mathrm{O}$, or with the modified one (6.2), (6.2′) we can simply get a reduced dimensioned model

$$Y_{1t} = F_{1t}'\theta_t + \nu_{1t}, \qquad \nu_{1t} \sim \mathrm{N}[0, V_{1t}],$$

$$\theta_t = G_t\theta_{t-1} + \omega_t, \qquad \omega_t \sim \mathrm{N}[0, W_t].$$

As it was pointed out in [10] the dimensionality of the observation vector varies over time, but this does not affect the validity of the Kalman filter.

Thus this method is powerful, if and only if the observational variance matrix is known. Even in this special case the modeller has to rearrange the model such that the above partition is obtained. Of course this in theory is always possible. However, in practice when the dimension of $\boldsymbol{Y}_t$ is large, this partitioning may involve modelling problems. This is when several subvectors of $\boldsymbol{Y}_t$ with different dimensions are missing over t.

For the rest of this chapter we leave the assumption of known observational variances and examine several dynamic models with unknown covariance structure. First we discuss the problems related with the use of the Wishart distribution in the models. These comments are not restricted to dynamic models. They apply to any normal regression models that make use of this distribution.

The use of the inverse Wishart distribution under the CCM, introduces certain limitations, related with missing observations. These have to do with the scalar degrees of freedom of the Wishart distribution. For example let $\boldsymbol{Y}_t = (Y_{1t}, Y_{2t})'$ be the observation vector and suppose that Y_{1t} is observed while Y_{2t} is missing. For the corresponding observational variance matrix, $(\Sigma | D_t) \sim \mathrm{W}_{n_t}^{-1}[\boldsymbol{S}_t]$, where $\boldsymbol{S}_t$, n_t are as defined in Theorem 2.4 of Section 2.3.2. Also write $\Sigma = \{\sigma_{ij}\}$, $\boldsymbol{S}_t = \{s_{ij}\}$, and $\boldsymbol{e}_t = (e_{1t}, e_{2t})'$ $(i, j = 1, 2)$. Then given D_t, the estimate of Σ, $\boldsymbol{S}_t$, may be derived recurrently according to

$$\begin{aligned} n_t s_{ij,t} &= n_{t-1} s_{ij,t-1} + e_{it} e_{jt} / Q_t, \\ n_t &= n_{t-1} + 1 \end{aligned}$$

for $i, j = 1, 2$ and Q_t as defined in Section 2.3.2.

Now since Y_{2t} is not available, neither is e_{2t} and so the calculation of

$s_{12,t}$, $s_{22,t}$ is not possible. It seems that one has to treat Y_{2t} as if it had been observed and use a value for it. Then, the updating $n_t = n_{t-1} + 1$ would be alright, but no such choice of Y_{2t} can be made. If Y_{2t} is treated as missing, then one can set $n_t = n_{t-1}$, but in this treats the observed value Y_{1t} as missing. It follows that with the CCM or with any multivariate model using the Wishart distribution, the only possible missing observation treatment is obtained when all the values of the observation vector at a specific point of time are missing.

Practitioners often assume that the unknown variance matrix $\boldsymbol{V}_t = \Sigma$ is diagonal. In such cases if $\Sigma = \text{diag}\{\sigma_{11}, \ldots, \sigma_{rr}\}$, then every σ_{ii} is modelled as having an inverse Gamma distribution, but with different degrees of freedom. Then the multivariate DLM factorises into several (r) univariate DLMs, and only then is missing observation analysis possible. However, conjugacy is lost and it is more sensible to consider from the beginning several independent univariate DLMs. In addition the assumption that the variance matrix is diagonal is often inappropriate.

In [40] a methodology was proposed for every multivariate DLM that provides full missing observation analysis and is based on the generalized inverse Wishart and T distributions, developed in Chapter 4. The following section presents this and extends the analysis for the DLMs of Chapters 3, 5.

6.3 Missing Observations

Following [40] we assume that missing observations are randomly collected. Our approach is based on excluding any missing values of the calculation of the updating equations (state and forecast distributions) thus excluding the unknown influence of these unobserved variables. This approach is explained for univariate dynamic models in [51, chapter 11].

Assume that in any multivariate DLM we observe all the $r \times 1$ vectors $\boldsymbol{Y}_i$, $i = 1, \ldots, t-1$. At time t some observations are missing (a subvector of $\boldsymbol{Y}_t$, or the whole vector $\boldsymbol{Y}_t$). To distinguish the former from the latter case we have the following definition.

Definition 6.1. *A* **partial missing observation vector** *is said to be any strict subvector of the observation vector that is missing. If the whole observation vector is missing it is referred to as* **total missing observation vector**.

This definition assumes that scalars are treated as 1×1 vectors.

6.3.1 The ECCM and the GMDLM

The ECCM (Extended Common Components Model) and the GMDLM (General Multivariate DLM) with observational variances were defined and developed in Sections 4.4 and 5.4 respectively.

Considering first the ECCM, it is clear that in the case of a total missing vector we have

$$(\Theta_t, \Sigma | D_t) \sim \mathrm{NGW}^{-1}[\boldsymbol{m}_t, \boldsymbol{C}_t, \boldsymbol{S}_t, \boldsymbol{N}_t, m_t], \tag{6.3}$$

where $\boldsymbol{m}_t = \boldsymbol{a}_t$, $\boldsymbol{C}_t = \boldsymbol{R}_t$, $\boldsymbol{S}_t = \boldsymbol{S}_{t-1}$, $\boldsymbol{N}_t = \boldsymbol{N}_{t-1}$, $m_t = m_{t-1}$, since no information comes in at time t. Now define the $r \times r$ matrix $\boldsymbol{U}_t =$

$\text{diag}\{i_{1t}, \dots, i_{rt}\}$ with

$$i_{jt} = \begin{cases} 1 & \text{if } Y_{jt} \text{ is observed,} \\ 0 & \text{if } Y_{jt} \text{ is missing,} \end{cases}$$

for all $1 \leq j \leq r$, where $\boldsymbol{Y}_t = (Y_{1t}, \dots, Y_{rt})'$.

Equation (6.3) still remains with recurrences

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t \boldsymbol{e}_t' \boldsymbol{U}_t, \tag{6.4}$$

$$\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{A}_t' Q_t u_t, \tag{6.5}$$

$$\boldsymbol{N}_t = \boldsymbol{N}_{t-1} + \boldsymbol{U}_t, \tag{6.6}$$

$$\boldsymbol{N}_t^{1/2} \boldsymbol{S}_t \boldsymbol{N}_t^{1/2} = \boldsymbol{N}_{t-1}^{1/2} \boldsymbol{S}_{t-1} \boldsymbol{N}_{t-1}^{1/2} + \boldsymbol{U}_t \boldsymbol{e}_t \boldsymbol{e}_t' \boldsymbol{U}_t / Q_t, \tag{6.7}$$

where $u_t = \dfrac{\text{trace}(\boldsymbol{U}_t)}{r}$. Some explanation for the above formulae is in order.

First note that if no missing observation occurs $\boldsymbol{U}_t = \boldsymbol{I}$, $u_t = 1$ and we have the standard recurrences as proved in Theorem 4.5. At the other extreme (total missing vector), $\boldsymbol{U}_t = \mathbf{O}$, $u_t = 0$ and we have equation (6.3). Consider now the case of partial missing observations. Equation (6.6) is the natural extension of the single degrees of freedom updating, see [51, page 351]. For equation (6.4) note that the zero's of the main diagonal of $\boldsymbol{U}_t$ convey the idea that the corresponding missing values elements of $\boldsymbol{m}_t$ remain unchanged and equal to $\boldsymbol{a}_t$. For example, consider the case of $r = 2$, $n = 2$ and assume that you observe Y_{1t}, but Y_{2t} is missing. Then

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \begin{pmatrix} A_{1t}(Y_{1t} - f_{1t}) & 0 \\ A_{2t}(Y_{1t} - f_{1t}) & 0 \end{pmatrix},$$

where $\boldsymbol{A}_t = (A_{1t}, A_{2t})'$. The zero's on the right hand side reveal that the 2nd column of $\boldsymbol{m}_t$ is the same as the 2nd column of $\boldsymbol{a}_t$. Similar comments apply for equations (6.5) and (6.7).

The proportion $(1-u_t)100\%$ is a measure of the influence of the missing observations at time t.

Suppose, now, that $\boldsymbol{Y}_t$ is an $r \times s$ matrix. Definition 6.1 can be extended to the following.

Definition 6.2. *A* **partial missing observation matrix** *is said to be any strict submatrix of the observation matrix that is missing. If the whole observation matrix is missing it is referred to as* **total missing observation matrix.**

This definition assumes that scalars are treated as 1×1 matrices.

Define $\boldsymbol{U}_{kt}$ to be the diagonal matrix $\boldsymbol{U}_{kt} = \text{diag}\{i_{1k,t}, \dots, i_{rk,t}\}$ with

$$i_{jk,t} = \begin{cases} 1 & \text{if } Y_{jk,t} \text{ is observed,} \\ 0 & \text{if } Y_{jk,t} \text{ is missing,} \end{cases}$$

where $\boldsymbol{Y}_t = \{Y_{jk,t}\}$, $(j = 1, \dots, r; k = 1, \dots, s)$.

Then, considering model (4.20), (4.20′), (4.21) (defined in page 80), the moments of equation (6.3) are updated via

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t \boldsymbol{e}_t' \prod_{k=1}^{s} \boldsymbol{U}_{kt}$$

$$\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{Q}_t \boldsymbol{A}_t' u_t$$

$$\boldsymbol{N}_t = \boldsymbol{N}_{t-1} + \sum_{k=1}^{s} \boldsymbol{U}_{kt}$$

$$\boldsymbol{N}_t^{1/2} \boldsymbol{S}_t \boldsymbol{N}_t^{1/2} = \boldsymbol{N}_{t-1}^{1/2} \boldsymbol{S}_{t-1} \boldsymbol{N}_{t-1}^{1/2} + \left(\prod_{k=1}^{s} \boldsymbol{U}_{kt} \right) \boldsymbol{e}_t \boldsymbol{Q}_t^{-1} \boldsymbol{e}_t' \left(\prod_{k=1}^{s} \boldsymbol{U}_{kt} \right),$$

where $u_t = \text{trace}(\prod_{k=1}^{s} \boldsymbol{U}_{kt})/r$.

Similar comments as in the case of $s = 1$ apply. This model may find an application to systematically missing data, e.g. an entire row of the data is missing. If the second row, for example, is always missing while the remaining rows are observed for all t, then $\boldsymbol{U}_{kt} = \text{diag}\{1, 0, 1, \ldots, 1\}$, for all $k = 1, \ldots, s$. For more details on this see [40].

As it was pointed out in the above reference the above methodology is not restricted to time series. Indeed normal theory together with the results of Sections 4.2, 4.3 provide the toolkit for missing observation analysis of Bayesian regression.

To illustrate this argument, suppose that an $r \times s$ observation matrix, $\boldsymbol{Y}$, is modelled such that

$$(\boldsymbol{Y}'|\Theta, \Sigma) \sim \text{N}[\boldsymbol{F}'\Theta, \boldsymbol{V}, \Sigma],$$
$$(\Theta, \Sigma) \sim \text{NGW}^{-1}[\boldsymbol{a}, \boldsymbol{R}, \boldsymbol{S}, \boldsymbol{N}, m],$$

where Θ is an appropriately defined state matrix or vector, and the quantities $\boldsymbol{F}, \boldsymbol{V}, \boldsymbol{a}, \boldsymbol{R}, \boldsymbol{S}, \boldsymbol{N}$ are suitably specified and known, while $\boldsymbol{\Sigma}$ is unknown.

Then, the above methodology implies that

$$\boldsymbol{Y}' \sim \text{GT}[\boldsymbol{f}', \boldsymbol{Q}, \boldsymbol{S}, \boldsymbol{N}, p],$$
$$(\Theta, \Sigma|\boldsymbol{Y}) \sim \text{NGW}^{-1}[\boldsymbol{m}, \boldsymbol{C}, \boldsymbol{S}^*, \boldsymbol{N}^*, m^*],$$

where

$$\begin{aligned} \boldsymbol{f}' &= \boldsymbol{F}'\boldsymbol{a}, & \boldsymbol{Q} &= \boldsymbol{F}'\boldsymbol{R}\boldsymbol{F} + \boldsymbol{S} \\ \boldsymbol{m} &= \boldsymbol{a} + \boldsymbol{A}\boldsymbol{e}'\prod_{k=1}^{s}\boldsymbol{U}_k, & \boldsymbol{C} &= \boldsymbol{R} - \boldsymbol{A}\boldsymbol{Q}\boldsymbol{A}'u \\ \boldsymbol{N}^* &= \boldsymbol{N} + \sum_{k=1}^{s}\boldsymbol{U}_k, & \boldsymbol{A} &= \boldsymbol{R}\boldsymbol{F}\boldsymbol{Q}^{-1}, \end{aligned}$$

$$N^{*1/2}S^*N^{*1/2} = N^{1/2}SN^{1/2} + \left(\prod_{k=1}^{s} U_k\right) eQ^{-1}e' \left(\prod_{k=1}^{s} U_k\right),$$

$$e = Y - f,$$

where $U_k = \mathrm{diag}\{i_{1k}, \ldots, i_{rk}\}$ with

$$i_{jk} = \begin{cases} 1 & \text{if } Y_{jk} \text{ is observed,} \\ 0 & \text{if } Y_{jk} \text{ is missing} \end{cases}$$

and $Y = \{Y_{jk}\}$ $(j = 1, \ldots, r; k = 1, \ldots, s)$, $u = \mathrm{trace}\left(\prod_{k=1}^{m} U_k\right)/r$.

Now coming back to the vector case, consider the normal GMDLM as developed in Section 4.7. The following distributions hold.

$$(\Sigma|D_t) \sim \mathrm{GW}^{-1}[S_t, N_t, m_t],$$

$$(\theta_t|\Sigma = S_0, D_t) \sim \mathrm{N}[m_t, C_t],$$

where

$$m_t = a_t + A_t^* A_t U_t e_t,$$

$$C_t = R_t - A_t^* S_0^{1/2} A_t U_t Q_t^* U_t A_t' S_0^{1/2} A_t^{*\prime},$$

$$N_t = N_{t-1} + U_t,$$

$$N_t^{1/2} S_t N_t^{1/2} = N_{t-1}^{1/2} S_{t-1} N_{t-1}^{1/2} + h_t^* h_t^{*\prime},$$

$$h_t^* = S_{t-1}^{1/2} Q_t^{-1/2} U_t e_t.$$

Similar comments apply as in the ECCM case.

Finally for completion purposes we consider the General Multivariate DLM with known variances. A discussion of this model appears in Chapter 2 (page 16) of this thesis and in [51, chapter 16]. Using the above definition of the matrix U_t the posterior distribution of θ_t at time t given D_t is

$$(\theta_t|D_t) \sim \mathrm{N}[m_t, C_t],$$

where

$$m_t = a_t + A_t U_t e_t,$$

$$C_t = R_t - A_t U_t Q_t U_t A_t'.$$

6.3.2 The Regression DLM

In Chapter 3 a multivariate regression DLM was developed, without any particular assumption of the unknown variance matrix Σ, (Section 3.3). Using a scalar degrees of freedom, its relationship with the CCM was examined in Section 3.4. However, as it has become evident so far, it is necessary to incorporate a matrix of degrees of freedom to the model in order to deal with partial missing observations. So instead of the estimate

$$n_t S_t = n_{t-1} S_{t-1} + r_t e_t',$$

with components as defined in Section 3.2, the proposed estimate is

$$N_t^{1/2} S_t N_t^{1/2} = N_{t-1}^{1/2} S_{t-1} N_{t-1}^{1/2} + r_t e_t',$$

where $N_t = \text{diag}\{n_{1t}, \ldots, n_{rt}\}$. It is clear that this estimate cannot be derived with a similar methodology to that in Chapter 3. The justification is due to its close relationship with the ECCM. Indeed, it is a matter of some simple algebra to see that the estimates of both models coincide.

After the above modifications the posterior distribution of $(\Theta_t | D_t)$ of Theorem 3.4 is

$$(\Theta_t | D_t, \Sigma = S_t) \sim \text{N}[m_t, C_t, S_t],$$

with updatings

$$m_t = m_{t-1} + A_t e_t' U_t,$$

$$C_t = \frac{1}{\delta}\left[I - \frac{C_{t-1} F_t F_t'}{\delta + F_t' C_{t-1} F_t} u_t \right] C_{t-1},$$

$$N_t = N_{t-1} + U_t,$$

$$N_t^{1/2} S_t N_t^{1/2} = N_{t-1}^{1/2} S_{t-1} N_{t-1}^{1/2} + U_t r_t e_t' U_t,$$

where the quantities A_t, e_t, r_t are defined in Chapter 3 and the remainder of this chapter. Notice that in the case of a total missing observation at time t, $C_t = R_t = C_{t-1}/\delta$, while in the case of a partial missing observation C_t depends upon u_t or depends on the proportion of the scalar missing values in Y_t.

This analysis is trivially extended to the General Regression DLM, see Section 3.6.

6.3.3 The General DLM of Chapter 5

This multivariate DLM was developed in Section 5.3, using the weak prior posterior modelling assumption. In Theorem 5.2 it was proved that the estimate of the unknown variance Σ given D_t is

$$n_t S_t = n_{t-1} S_{t-1} + S_{t-1}^{1/2} Q_t^{-1/2} e_t e_t' Q_t^{-1/2} S_{t-1}^{1/2},$$

where n_t are the scalar degrees of freedom, and the remaining components are as defined in Chapter 5. Again we face the need for upgrading the scalar degrees of freedom to the matrix form. One possible way to upgrade the above estimate using a matrix of degrees of freedom is

$$N_t^{1/2} S_t N_t^{1/2} = N_{t-1}^{1/2} S_{t-1} N_{t-1}^{1/2} + S_{t-1}^{1/2} Q_t^{-1/2} e_t e_t' Q_t^{-1/2} S_{t-1}^{1/2}, \qquad (6.8)$$

where $N_t = \text{diag}\{n_{1t}, \dots, n_{rt}\}$. However, at the time of writing there is no procedure to support this, except the similarity with the other models.

Theorem 5.2 may be called upon to provide a solution. By redefining $\mathcal{A}_t = N_t^{-1/2} S_{t-1}^{1/2} Q_t^{-1/2}$, and using the weak assumption

$$\text{vech}(\Sigma - \mathcal{A}_t e_t e_t' \mathcal{A}_t') \perp_1 Y_t | D_{t-1} \tag{6.9}$$

it is easily shown that

$$N_t^{1/2} S_t N_t^{1/2} = N_t^{1/2} S_{t-1} N_t^{1/2} - S_{t-1} + S_{t-1}^{1/2} Q_t^{-1/2} e_t e_t' Q_t^{-1/2} S_{t-1}^{1/2}. \tag{6.10}$$

Equations (6.8) and (6.10) coincide if $N_t = \text{diag}\{n_t, \dots, n_t\}$ and give the estimate of Theorem 5.2. In general the diagonal elements of the matrix S_t in each of these equations are always the same and the only difference lies on the off-diagonal elements. To see this denote the matrix $S_{t-1} = \{s_{ij,t-1}\}$, $(i, j = 1, \dots, r)$ and $N_t = \text{diag}\{n_{1t}, \dots, n_{rt}\}$ with the usual updating $n_{it} = n_{i,t-1} + 1$, $(i = 1, \dots, r)$. Now equation (6.8) implies

$$N_{t-1}^{1/2} S_{t-1} N_{t-1}^{1/2} = \left\{ n_{i,t-1}^{1/2} n_{j,t-1}^{1/2} s_{ij,t-1} \right\},$$

while equation (6.10)

$$N_t^{1/2} S_{t-1} N_t^{1/2} - S_{t-1} = \left\{ (n_{it}^{1/2} n_{jt}^{1/2} - 1) s_{ij,t-1} \right\}.$$

for all $i, j = 1, \dots, r$.

From the above equations it is clear that $N_{t-1}^{1/2} S_{t-1} N_{t-1}^{1/2} \neq N_t^{1/2} S_{t-1} N_t^{1/2} - S_{t-1}$, hence the two estimates are in general different. Our view is in favour of the estimate produced by equation (6.10). Note that this equation can also take the form

$$S_t = S_{t-1} + N_t^{-1/2} S_{t-1}^{1/2} (Q_t^{-1/2} e_t e_t' Q_t^{-1/2} - I) S_{t-1}^{1/2} N_t^{-1/2}, \tag{6.11}$$

with the usual updating $\boldsymbol{N}_t = \boldsymbol{N}_{t-1} + \boldsymbol{I}$, when there is no missing observation.

Using assumptions (6.9), (6.10) Theorem 5.3 is updated to incorporate $\boldsymbol{N}_t$. Then the quantities $\boldsymbol{m}_t$, $\boldsymbol{a}_t$, $\boldsymbol{f}_t$, $\boldsymbol{C}_t$, $\boldsymbol{R}_t$, $\boldsymbol{Q}_t$ of Theorem 5.3 is as in this theorem just replacing $\boldsymbol{S}_t$ by equation (6.11).

Moving now to the missing observation analysis, if any subvector of $\boldsymbol{Y}_t$ (including possibly the whole $\boldsymbol{Y}_t$) is missing, the quantities $\boldsymbol{m}_t$, $\boldsymbol{C}_t$, $\boldsymbol{N}_t$, $\boldsymbol{S}_t$ are updated by

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t\boldsymbol{U}_t\boldsymbol{e}_t,$$

$$\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{U}_t\boldsymbol{Q}_t\boldsymbol{U}_t\boldsymbol{A}_t',$$

$$\boldsymbol{N}_t = \boldsymbol{N}_{t-1} + \boldsymbol{U}_t,$$

$$\boldsymbol{S}_t = \boldsymbol{S}_{t-1} + \boldsymbol{N}_t^{-1/2}\boldsymbol{S}_{t-1}^{1/2}(\boldsymbol{Q}_t^{-1/2}\boldsymbol{U}_t\boldsymbol{e}_t\boldsymbol{e}_t'\boldsymbol{U}_t\boldsymbol{Q}_t^{-1/2} - \boldsymbol{U}_t)\boldsymbol{S}_{t-1}^{1/2}\boldsymbol{N}_t^{-1/2}.$$

In Section 2.7 it was mentioned that a simple, but effective, form of intervention is to treat outliers, or in general, observations that are not to be used for future updating, as missing. Section 6.3 gives a complete contribution to this form of intervention allowing for a treatment of a subvector of $\boldsymbol{Y}_t$ as an outlier. For example, if $\boldsymbol{Y}_t = (Y_{1t}, Y_{2t})'$ and the scalar Y_{1t} is an outlier at t, but Y_{2t} is not, then the above methodology will ignore only Y_{1t} and the forecasts of $Y_{1,t+1}$ will be based on the previous $Y_{1,t-1}$, while the forecast of $Y_{2,t+1}$ will be based on Y_{2t}.

The above methodology provides a powerful analysis for every multivariate DLM with respect to any missing observations.

The retrospective and filtering distributions of the relevant models of Sections 4.5, and 5.4, are easily upgraded in the presence of any partial or total missing observation vector. Indeed, it is obvious that under the modifications of $\boldsymbol{m}_t$, $\boldsymbol{C}_t$, $\boldsymbol{S}_t$, and $\boldsymbol{N}_t$ of the relevant Sections 6.3.1, 6.3.3, Theorems 4.8 and 5.5 still hold. Also, Theorem 4.10 (theorem of deleting observations for the ECCM) is valid with the following modifications

$$\boldsymbol{a}_{t,k} = \boldsymbol{a}_t(-k) - \boldsymbol{A}_t(-k)\boldsymbol{e}_t'(-k)\boldsymbol{U}_t,$$

$$\boldsymbol{R}_{t,k} = \boldsymbol{R}_t(-k) + \boldsymbol{A}_t(-k)Q_t(-k)\boldsymbol{A}_t'(-k)u_t,$$

$$\boldsymbol{N}^{1/2}\boldsymbol{S}_{t,k}\boldsymbol{N}^{1/2} = \boldsymbol{N}_t^{1/2}\boldsymbol{S}_t\boldsymbol{N}_t^{1/2} - \boldsymbol{U}_t\boldsymbol{e}_t(-k)\{Q_t(-k)\}^{-1}\boldsymbol{e}_t'(-k)\boldsymbol{U}_t,$$

$$\boldsymbol{N} = \boldsymbol{N}_t - \boldsymbol{U}_t,$$

where the $r \times r$ matrix $\boldsymbol{U}_t$ is as defined in Section 6.3.1.

The extension to deleting a subvector of $\boldsymbol{Y}_t$ is straight-forward in the light of Theorem 4.11 and the above modifications. In such a way the concept of deleting any observations and missing observations can be put together, under the framework of the ECCM. For example, an application may be the deletion of an observation vector containing missing observations.

The development of Chapter 4 was done to enable current models to handle partial missing observation problems. Here, the necessary modifications of the moments of the relevant distributions for the reference analysis (Section 4.6) are presented.

If there are missing observations the distributions of the theorems of Section 4.6 remain improper at $t = t_{n,r} = n + (r+1)/2$. More observations are needed to overlap the gap of the missing values. Let m denote the total missing single observations and $t_{n,r,m}$ the minimum time after which the distributions are proper. If $m = kr$, for some $k \in \mathbb{N}^*$, we set $t_{n,r,m} = n + (r+1)/2 + k$. If $m \neq kr$, for all $k \in \mathbb{N}^*$, then there exist positive integers λ, ν such that $m = \lambda r + \nu$ with $0 < \nu < r$. In this case we set $t_{n,r,m} = n + (r+1)/2 + \lambda + 1$. Thus in all theorems of Section 4.6, $t_{n,r}$ is replaced by $t_{n,r,m}$. Of course if

$m = 0$ we just set $t_{n,r,m} = t_{n,r}$.

Since the distributions of Theorem 4.12 are proper for $t \geq t_{n,r,m}$ the updatings of $\boldsymbol{m}_t$, $\boldsymbol{C}_t$, $\boldsymbol{S}_t$, $\boldsymbol{N}_t$ follow the respective of Section 6.3.1 for the ECCM. Similar comments apply for Theorems 4.13, 4.14, and 4.15.

Table C.2, on page 244, shows the essential calculus of missing and observed values that motivated all the above analysis.

6.4 Unequally Spaced Observations

In this thesis so far it has been assumed that the observations are recorded for each integer value of t, normally $t, t+1, \ldots$ In this section we deal with the problem of observations at any general space time sequence $t+k_1, t+k_2, \ldots, t+k_n$, where k_i, $(i = 1, \ldots, n)$, are any integers such that $k_1 < \cdots < k_n$, for $n > 1$.

It is worthwhile noting that this causes major problems for ARIMA models. However, under the Bayesian framework this problem is easily resolved. The idea is to create a new partition of the interval $[t+k_1, t+k_n]$ by adding new points to the existing $t+k_1, \ldots, t+k_n$ such that the new partition consists of equally spaced points. Then these observations that correspond to the adding points are treated as missing. In the next paragraph we present a simple algorithm of the above methodology.

First define d_i to be the distance of the neighbourhood points $t+k_{i-1}$ and $t+k_i$, $(i = 2, \ldots, n)$. Further let d be the minimum of all these distances, so

$$d = \min\{d_i, i = 2, \ldots, n\} = \min\{k_i - k_{i-1}, i = 2, \ldots, n\}.$$

If $d_i = d$, for all i, we have an equally spaced series. If there exists at least one j, $(2 \leq j \leq n)$, such that $d_j > d$, then the following algorithm may be used

Step 1: Calculate d_i, for all $i = 2, \ldots, n$, and then $d = \min\{d_i, i = 2, \ldots, n\}$.

Step 2: For j such that $d_j > d$, calculate the distance $d^{(j)} = d_j - d$ and add any observation $\boldsymbol{Y}^{(j)}$, $2 \leq j \leq n$.

(i) If $d^{(j)} > d$, then repeat step 2 for $d_j = d^{(j)}$;

(ii) If $d^{(j)} < d$, then set $d = d^{(j)}$ and repeat step 2;

(iii) If $d^{(j)} = d$, then go to the next j.

Step 3: Treat all new observations $\boldsymbol{Y}^{(j)}$ as missing.

For example assume that our series comprises 5 observations in the set $[1, 3, 6, 7, 15]$. According to Step 1, we calculate $d_2 = 2$, $d_3 = 3$, $d_4 = 1$, $d_5 = 8$ and $d = 1$. In Step 2, $d_2 > d$; we calculate $d^{(2)} = 1$ and we add the first observation, $\boldsymbol{Y}^{(2)}$. Then we note that $d^{(2)} = d$ and so we go to $j = 3$. This procedure is repeated until we construct an equally spaced partition of the set $[1, 15]$. The algorithm is proved very fast and effective, especially when the different d_i are quite similar.

Alternatively, the time index $t \in \mathbb{N}_p$ $(p > 1)$ may be changed to a subset of $\mathbb{N}_p$ comprising consecutive integers that correspond to the observed time series $\boldsymbol{Y}_t$. Then t is transformed to a new index t^* and consequently $\boldsymbol{Y}_t$ is transformed to a new series $\boldsymbol{Y}_{t^*}$. This method is computational efficient, but care must be drawn when forecasting for the actual series $\boldsymbol{Y}_t$ is desired. This finds application to commodity short term forecasting where observations are only collected in trading days and there is not particular interest for weekends or bank holidays. This is further explored through an example in the following section.

6.5 Illustration

In this section the multivariate first order polynomial model is used to model the changes of a London Metal Exchange aluminium price time series. The data are briefly described and then the model is built.

The London Metal Exchange (LME) is the world's premier non-ferrous metals market, with highly liquid contracts. Its trading customers may be metal industries or individuals (sellers or buyers).

LME's main functions are

(i) to provide a daily price for its metals which are relied upon worldwide industry;

(ii) to provide futures and traded option contracts that allow for prices to be locked in (this risk management function is known as hedging);

(iii) to act as a deliverer of last resort by authorizing warehouses to store approved brands of metal.

LMEX, the London Metal Exchange index, is a base metals index comprising the six primary non-ferrous metals traded on the Exchange: aluminium, copper grade A, standard lead, primary nickel, tin, and zinc. More details about the LME may be found on its website: www.lme.co.uk.

More than 460 brands of metal from 66 countries are approved as "good delivery" against LME contracts. Aluminium of Greece S.A.I.C., member of the Pechiney Group, is one of these brands.

In this example we concentrate on the aluminium official prices provided by Reuters. The data are collected for every trading day and they are taken from March 2000 to February 2001, see Table C.3 (page 245). Some explanations are in order. There are 4 main columns (Cash, 3 months, 15 months, 27 months), each of which comprises of two sub-columns (Bid price and Ask price) and the last column (Settimal) is an overall weighted index. The numbers shown on this table are the prices per tonne of aluminium. The Cash column is the daily/current closing price (Bid/Ask) of aluminium. The next 3 columns are the relevant future contracts (3 months, 15 months, 27 months).

The objective of this work is to produce an easy to implement forecasting system for such kind of data. The assumptions made are kept to a minimum. Our purpose is to build a simple yet effective forecasting system.

With a brief look at Table C.3, we see that the difference between the Bid and Ask prices is very small. Let $\boldsymbol{Y}_t = (Y_{1t}, \ldots, Y_{9t})'$ denotes the vector that summarises all the 9 aluminium prices at each time t. In other words Y_{1t} is the Cash Bid price at time t, Y_{2t} the Cash Ask price at time t, Y_{3t} the 3 months Bid price at time t, Y_{4t} the 3 months Ask price at time t, ..., Y_{9t} the Sett price at time t. The time t is measured in days starting from March 2000. So the last date, 22 February 2001, is the 359 day of the year March 2000 / February 2001. Table C.3 shows that data relate to unequally spaced days. This happens because on some days of each month (weekends plus bank holidays etc) there is no trading or data collection. For example such days are the 1,4,5,11,12,18,19,25,26/3/2000 and the 1,2,8,9,15,16,22,23,29,30/4/2000. Also we note that there are total missing observations on 21,24/4/2000, 29/5/2000, 28/8/2000, 25,26/12/2000. Figure 6.1 demonstrates the bid price time series for the variables Cash, 3 months, 15, months, and 27 months as shown from the top.

A usual "efficient market" practice for commodity price short term forecasting is to take as one step forecast just the current observed value. This, in theory, works well except when there are shocks that cannot be predicted with the current model. These shocks are clearly identified by modelling instead of the actual series $\boldsymbol{Y}_t$, the difference $\boldsymbol{Y}_t - \boldsymbol{Y}_{t-1}$. Suppose that at time $t-1$ the observation $\boldsymbol{Y}_{t-1}$ is available. According to the above discussion commodity short term forecasting suggests that $\boldsymbol{Y}_t$ is just predicted by $\boldsymbol{Y}_{t-1}$, unless there is evidence for a shock. In this case intervention may be called upon to provide a sensible forecast.

Table C.4 (page 258) shows the difference series $\boldsymbol{Z}_t = \boldsymbol{Y}_t - \boldsymbol{Y}_{t-1}$ for all the trading days. Weekends and bank holidays have been excluded such that the above difference is always sensible. That is there are 247 consecutive trading days from March 2000 to February 2001. Here and in Table C.4, for simplicity and demonstrative clarity, we deal with the first 6 months $(t = 1, \dots, 135)$.

A simple summary analysis showed that the maximum value among all variables is as high as 68, while the minimum -44.5. The maximum mean of each variable is 0.31, while the minimum mean is 0.007. The correlation

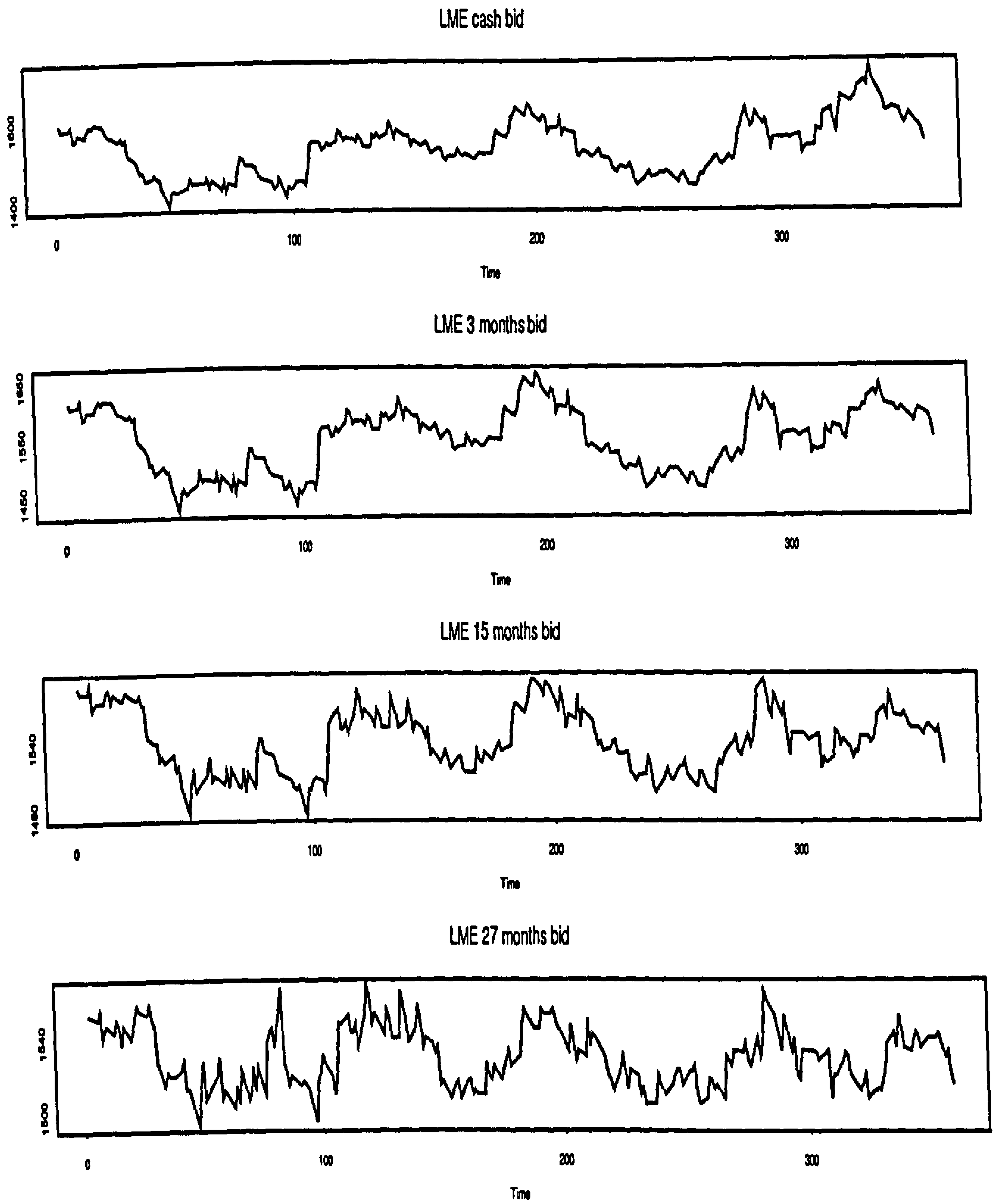

Figure 6.1: London Metal Exchange aluminium closing prices

matrix of the variables is

$$\begin{pmatrix} 1 & 0.999 & 0.987 & 0.987 & 0.903 & 0.903 & 0.694 & 0.694 & 0.999 \\ 0.999 & 1 & 0.986 & 0.987 & 0.905 & 0.905 & 0.695 & 0.695 & 0.999 \\ 0.987 & 0.986 & 1 & 0.999 & 0.932 & 0.932 & 0.729 & 0.729 & 0.987 \\ 0.987 & 0.987 & 0.999 & 1 & 0.932 & 0.932 & 0.728 & 0.728 & 0.987 \\ 0.903 & 0.904 & 0.932 & 0.932 & 1 & 1 & 0.853 & 0.853 & 0.904 \\ 0.903 & 0.904 & 0.932 & 0.932 & 1 & 1 & 0.853 & 0.853 & 0.904 \\ 0.694 & 0.695 & 0.729 & 0.728 & 0.853 & 0.853 & 1 & 1 & 0.695 \\ 0.694 & 0.695 & 0.729 & 0.728 & 0.853 & 0.853 & 1 & 1 & 0.695 \\ 0.999 & 0.999 & 0.987 & 0.987 & 0.904 & 0.904 & 0.695 & 0.695 & 1 \end{pmatrix}.$$

We note that most of the off-diagonal elements of the above matrix are very close to 1 that show the high correlation between the variables. For demonstrative clarity, the elements of this matrix are rounded about the first 3 digits. This may cause the confusion that the correlation matrix is singular. In practice much interest will be on the Ask variables, thus instead of dealing with 9 variables it is proposed that only 5 variables are considered (the 4 Ask sub-variables and the settimal one).

Now, given $\boldsymbol{Y}_{t-1}$ the difference $\boldsymbol{Y}_t - \boldsymbol{Y}_{t-1}$ is expected to have zero mean ($\boldsymbol{Y}_t$ is predicted as $\boldsymbol{Y}_{t-1}$) and some unknown variance. This is further supported by the summary analysis. The entire distribution will be a symmetric distribution. The histogram of the data for the Cash Bid variable is shown in Figure 6.2 and a normal probability plot for this variable in Figure 6.3. Both these figures show that an assumption of a normal distribution for the first difference is very considerable. Figure 6.2 shows that the density is not far from a normal one, and Figure 6.3 concludes the same since the values for the Cash Bid variable form approximately a straight line. Similar results

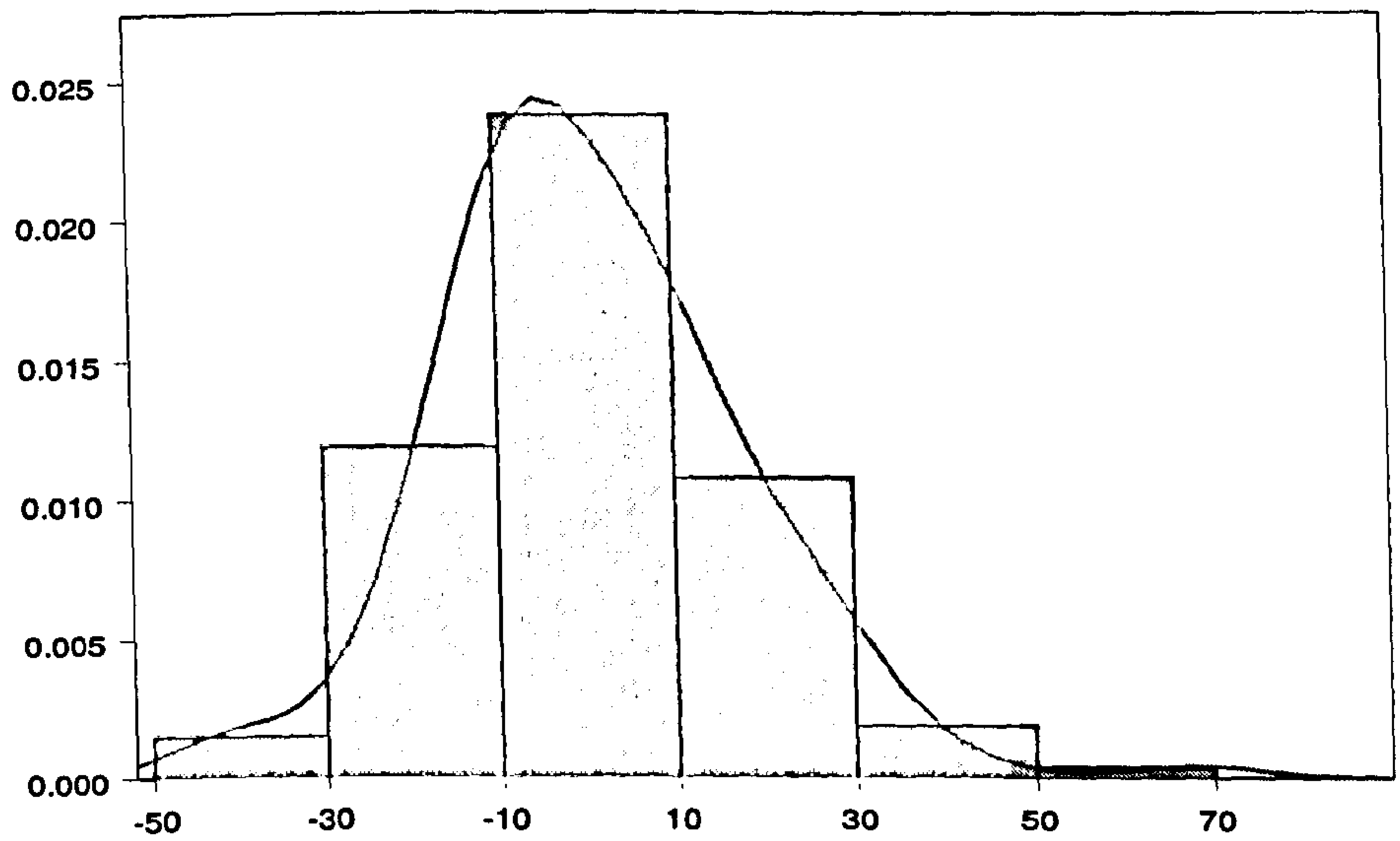

Figure 6.2: Histogram and density estimation of the Cash Bid time series

are obtained from the remaining variables.

Although the assumption of a normal distribution is supported by the above graphs, there is evidence for the presence of outliers. Influential observations like at $t = 64(-44)$ and at $t = 72(68)$ for the first variable, could yield a symmetric distribution with longer tails than a normal one. This means that if a normal distribution is considered the forecasts will not be very good for extreme observations, giving high one step errors. This is overcome by employing expert intervention techniques.

According to the above the first order polynomial normal DLM $\{\boldsymbol{I}_9, \boldsymbol{I}_9, \Sigma, \boldsymbol{W}_t\}$ is appropriate for modelling $\boldsymbol{Z}_t$. This model is given by

$$\boldsymbol{Z}_t = \boldsymbol{\mu}_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim \mathrm{N}[\boldsymbol{0}, \Sigma],$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathrm{N}[\boldsymbol{0}, \boldsymbol{W}_t],$$

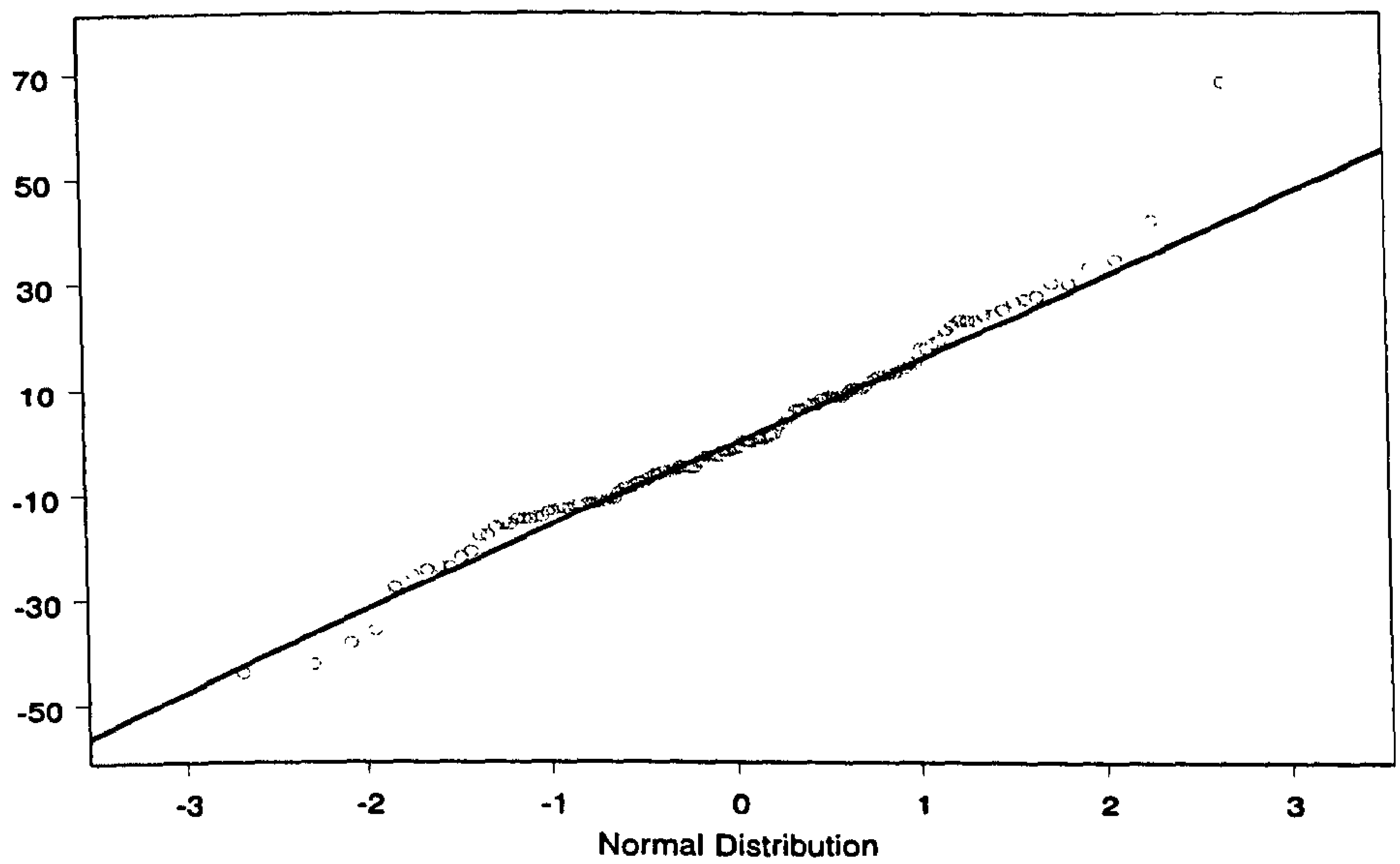

Figure 6.3: Normal probability plot for the Cash Bid time series

where $\boldsymbol{Z}_t$ is defined as above, $\boldsymbol{\mu}_t$ is the level of the series, $\boldsymbol{\Sigma}$ an unknown 9×9 variance matrix, and $\boldsymbol{W}_t$ an 9×9 known evolution matrix. As usual, D_t is the available information at time t and $\boldsymbol{W}_t$ is specified with a discount factor δ. All $Y_{1t}, \ldots, Y_{9t}$ have a similar behaviour throughout t, thus it is sensible to use a single discount factor.

For such kind of data it is expected that the elements of $\boldsymbol{W}_t$ will be small compared with those of $\boldsymbol{\Sigma}$ and so a fairly high discount factor $\delta = 0.9$ has been used.

Initially it is set

$$(\boldsymbol{\mu}_0 | D_0) \sim [\mathbf{0}, \boldsymbol{I}_9],$$

and $\boldsymbol{N}_0 = \boldsymbol{I}_9$, $\boldsymbol{S}_0 = \boldsymbol{I}_9$, where $\boldsymbol{N}_0$, $\boldsymbol{S}_0$ are as those used in Section 6.3.3.

Using assumption (6.9) of Section 6.3.3 and considering normality, Theorem 5.3 holds.

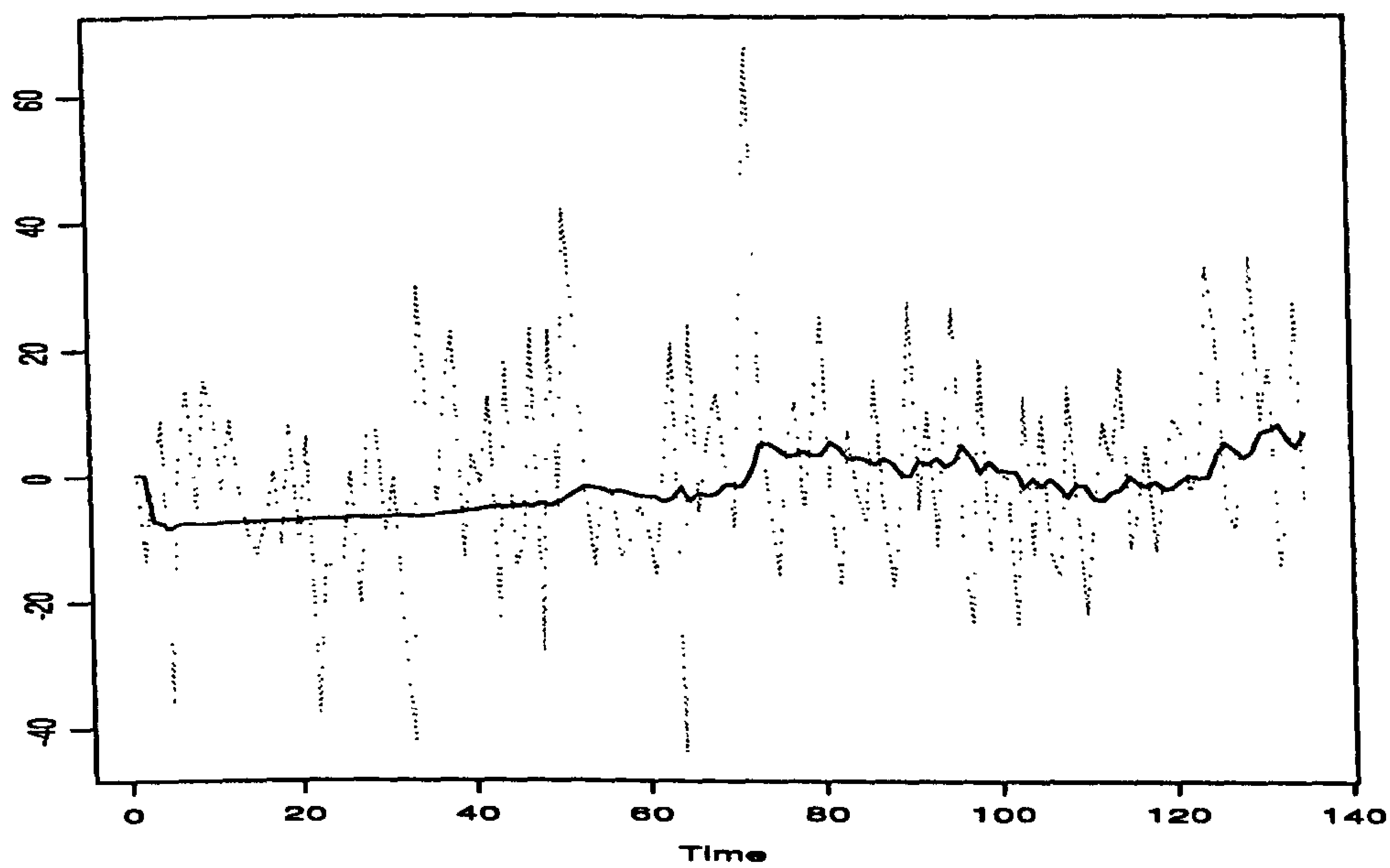

Figure 6.4: One step forecasts vs actual values for the difference time series

In this example we concentrate on $Z_{1t} = Y_{1t} - Y_{1,t-1}$. It must be noted that this model produces forecasts for all $Z_{it} = Y_{it} - Y_{i,t-1}$. The one step forecasts are plotted against the actual values Z_{1t} in Figure 6.4. The continuous line represents the forecasts, while the dotted line the real values. Figure 6.5 shows the one step forecast errors.

Both figures confirm what it was stated before: the model is good except when shocks appear. Outliers are responsible for high one step errors. The modeller has to decide upon the outlier limits beyond which a value will be detected as an outlier. Since the highest error is 67.57 ($t = 72$), the outlier limits were chosen as ± 30. This is confirmed by looking at Figures 6.3 and 6.5.

Thus outliers were detected at times $t = 22, 33, 34, 51, 64, 71, 72, 124, 129$. These outliers correspond to a proportion of 6.66% of the total observations.

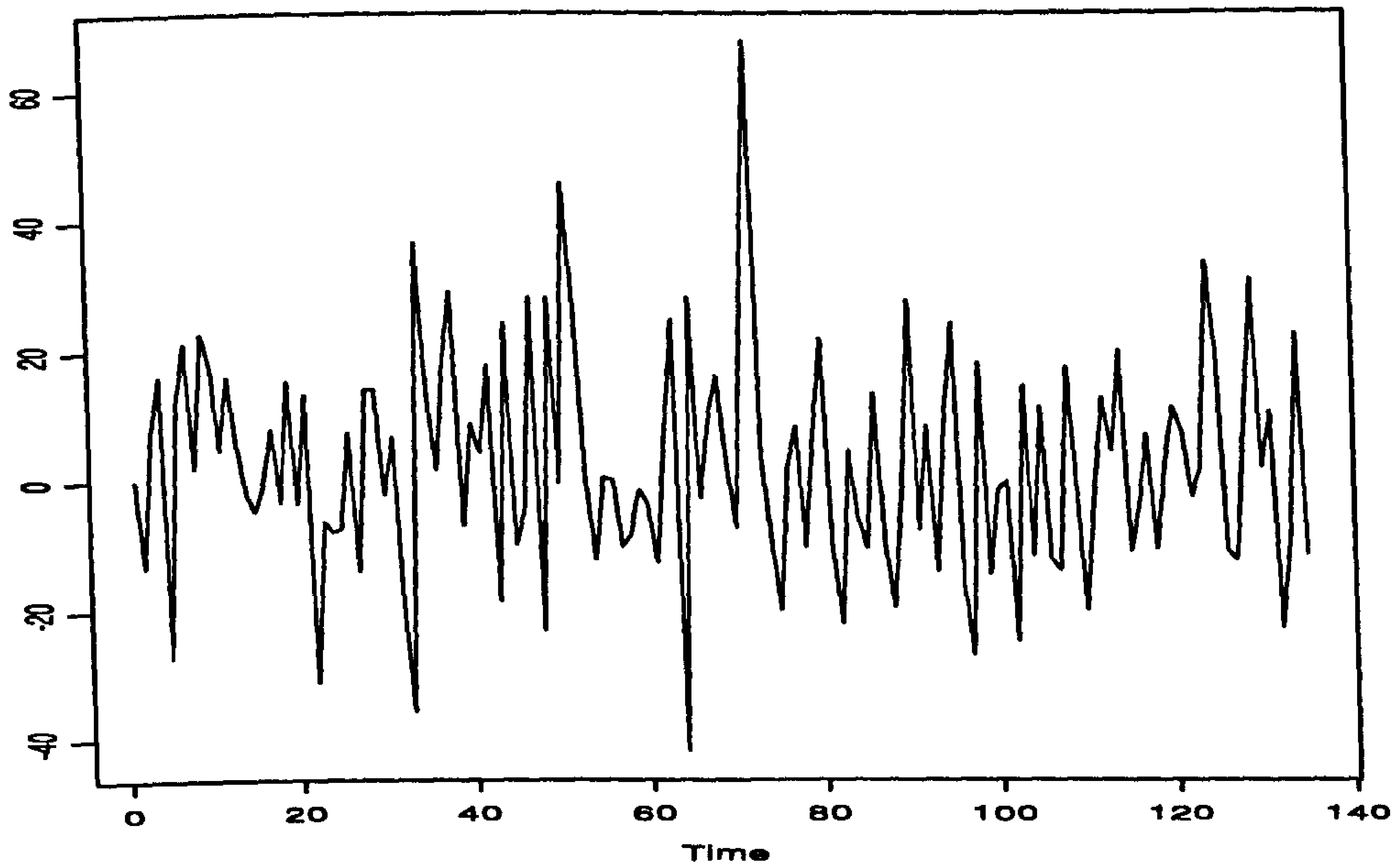

Figure 6.5: One step errors for the difference time series

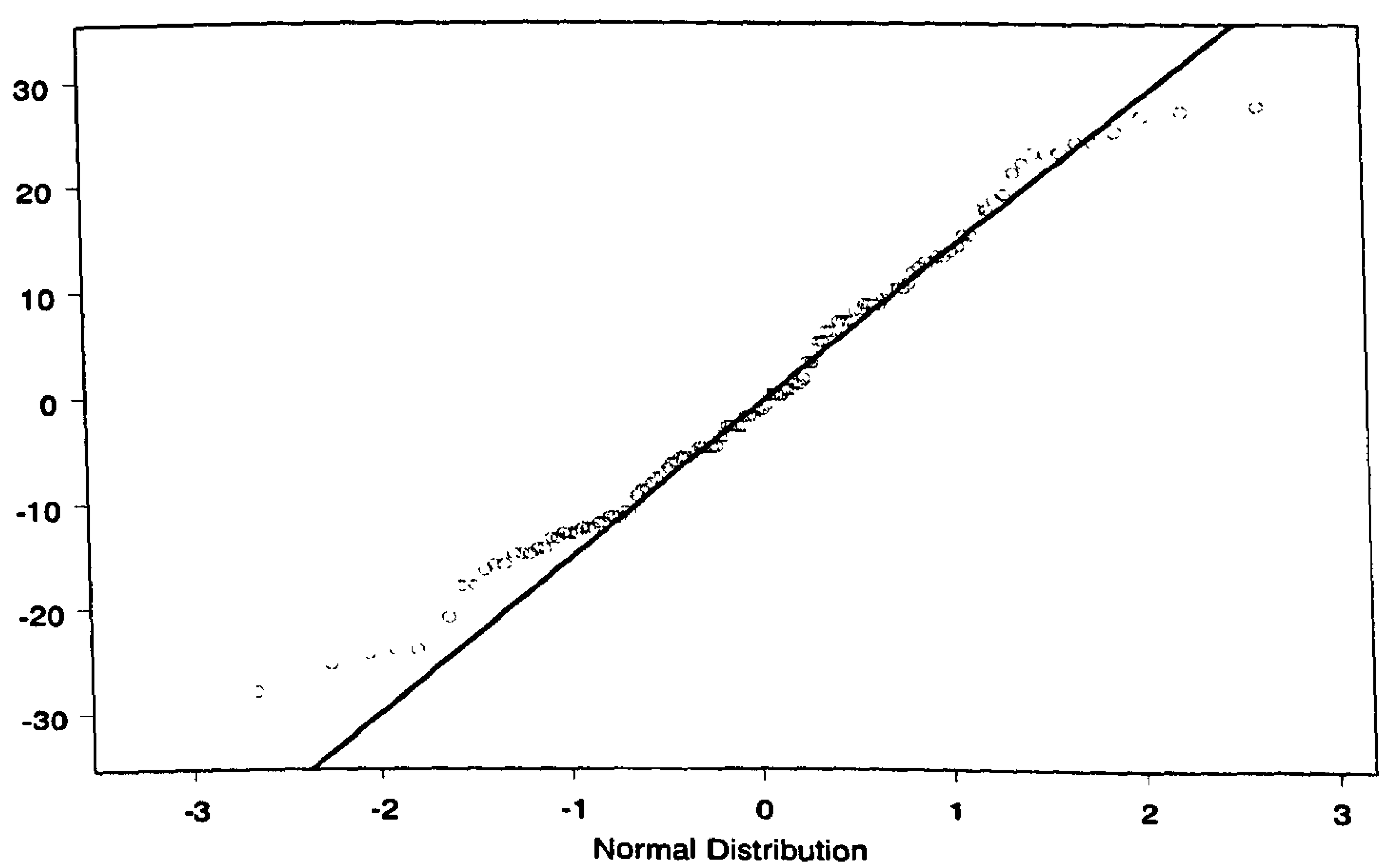

Figure 6.6: Normal probability plot excluding outliers

Figure 6.6 is the normal probability plot for the first variable (Cash Bid), if these outliers are excluded. We see that data are very close to a normal distribution. This is a simple method of outlier detection that identifies the most influential observations. More advanced methods may be considered like standardized residual plots. Here the goal is to show how the modeller can intervene after influential observations have been detected and not providing a thorough error analysis. With all these expert intervention is essential.

Before we proceed with intervention we clarify the statement that the values of Σ (or those of the estimate of it) will be very large compared with those of $\boldsymbol{W}_t$. Figure 6.7 compares the values of $s_{11,t}$ with $50w_{11,t}$, where $\boldsymbol{S}_t = \{s_{ij,t}\}$ (the estimate of Σ at t) and $\boldsymbol{W}_t = \{w_{ij,t}\}$, $(i, j = 1, \ldots, 9)$. The continuous line is for $s_{11,t}$, while the dotted line is for $50w_{11,t}$ We see that $s_{11,t}$ is huge compared with $w_{11,t}$.

A simple mode of intervention was discussed in Section 2.7. According to this, a noise term $\boldsymbol{\omega}_t^*$ is added to the evolution equation such that

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\omega}_t + \boldsymbol{\omega}_t^*,$$

where $\boldsymbol{\omega}_t^*$ is uncorrelated with $(\boldsymbol{\mu}_{t-1}|D_{t-1})$ and $\boldsymbol{\omega}_t^* \sim \mathrm{N}[\boldsymbol{w}_t^*, \boldsymbol{W}_t^*]$, for some known quantities $\boldsymbol{w}_t^*$ and $\boldsymbol{W}_t^*$.

Write $\boldsymbol{w}_t^* = (w_{1t}^*, \ldots, w_{9t}^*)'$. For all $t \neq 22, 33, 34, 51, 64, 71, 72, 124, 129$ we set $\boldsymbol{w}_t^* = \boldsymbol{0}$ and $\boldsymbol{W}_t^* = \boldsymbol{O}$ so that values that are not outliers are predicted by current forecasts. For intervening at Y_{1t} we set $w_{1,22}^* = -21.4$, $w_{1,33}^* = -20.8$, $w_{1,34}^* = 24.6$, $w_{1,51}^* = 30.8$, $w_{1,64}^* = -24$, $w_{1,71}^* = 21.8$, $w_{1,72}^* = 52.6$, $w_{1,124}^* = 23.3$, $w_{1,129}^* = 21.8$, and $\boldsymbol{W}_t^* = 100\boldsymbol{I}_9$, for all intervening points t. Similarly all the w_{it}^* can be set. The variance $\boldsymbol{W}_t^*$ reflects the uncertainty associated

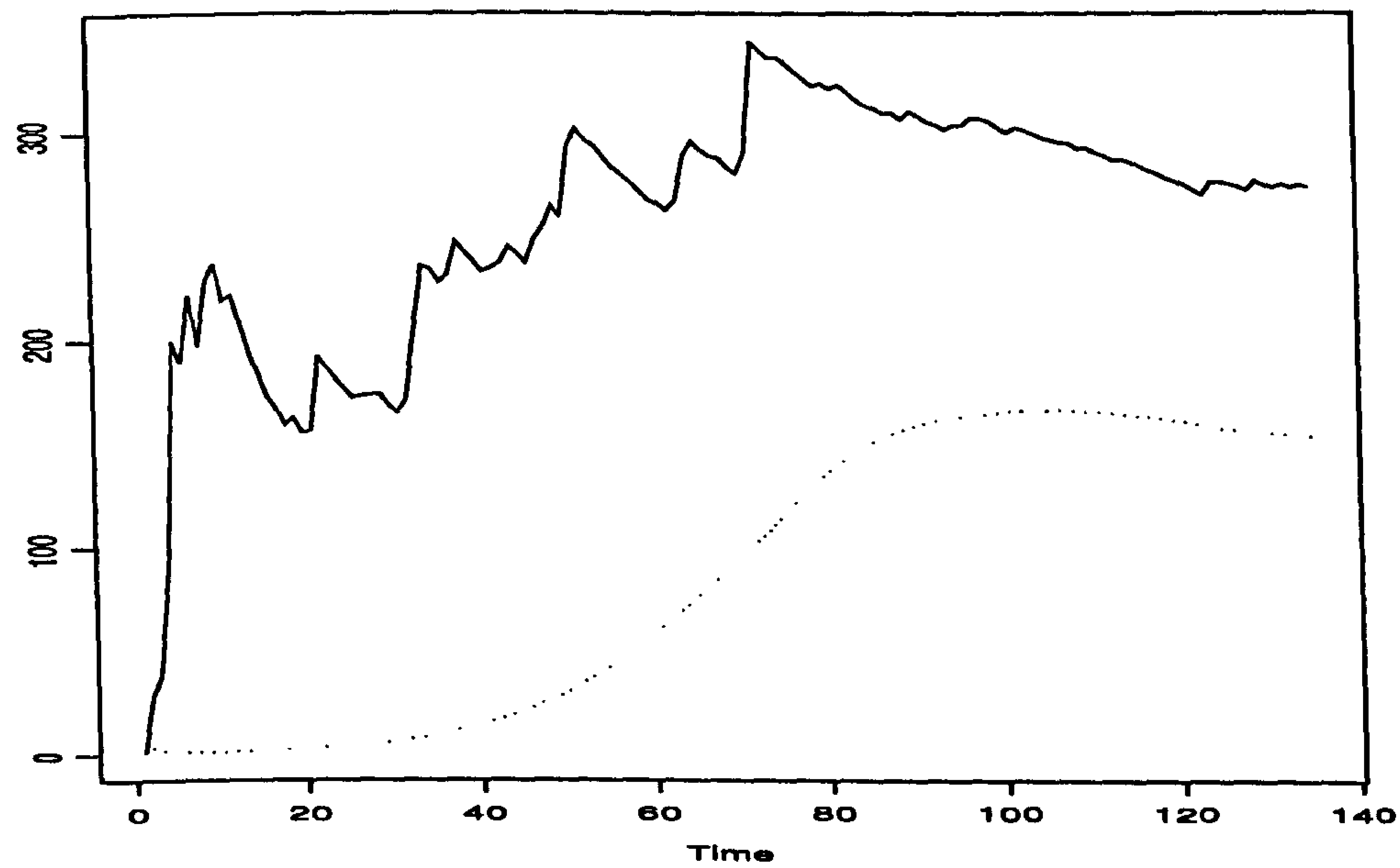

Figure 6.7: Comparison of the observational and evolution variances

with the expectations w_t^*.

Figure 6.8 shows the one step forecasts (continuous line) after intervention together with the actual values Z_{1t} (dotted line). We observe a much better prediction performance. This is more clear in Figure 6.9, where the associated errors are drawn. The maximum of the absolute value of the errors does not exceed 29.

As stated before the values of the estimate $\boldsymbol{S}_t$ are very high compared with those of $\boldsymbol{W}_t$. Intervention will not dramatically affect $\boldsymbol{S}_t$. However, we found a slight decline in all the elements of $\boldsymbol{S}_t$ after intervention. Here we

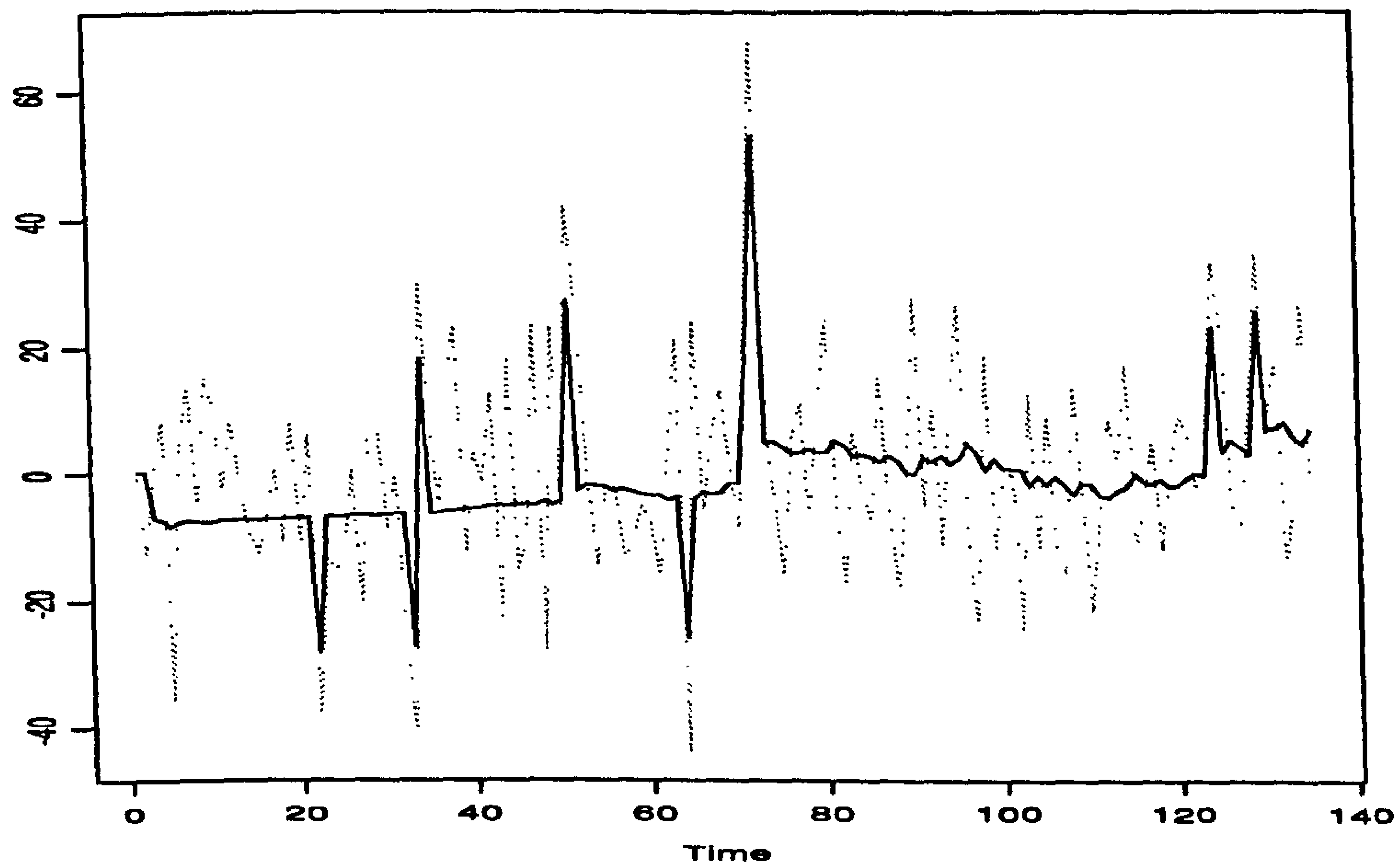

Figure 6.8: One step forecasts vs actual values for the difference time series after intervention

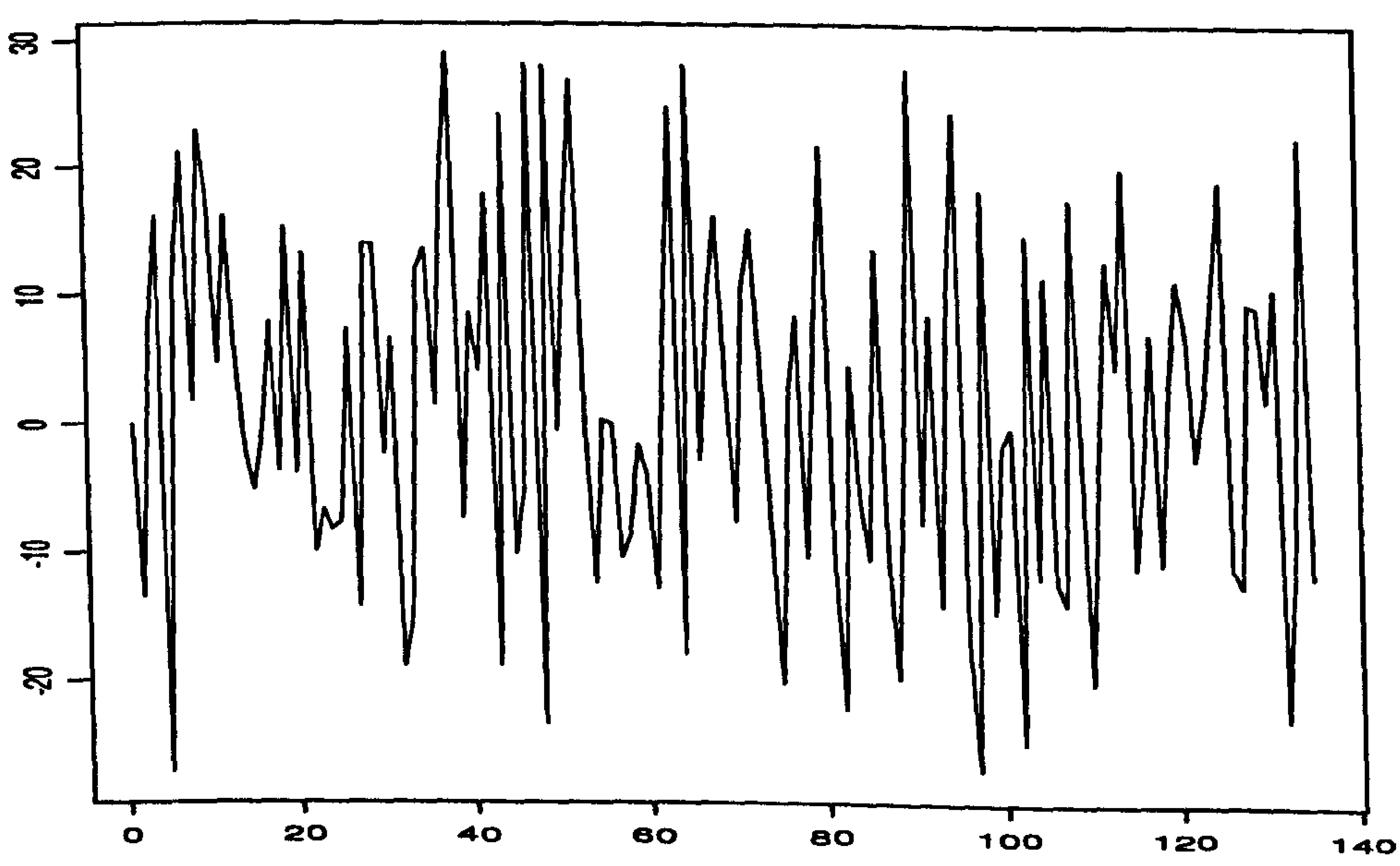

Figure 6.9: One step errors for the difference time series after intervention

have these estimates for times $t = 20, 40, 80, 120$. So

$$S_{20} = \begin{pmatrix} 155.3 & 152.3 & 117.1 & 118.7 & 65.1 & 65.1 & 36.8 & 36.8 & 152.3 \\ 152.3 & 149.5 & 114.8 & 116.4 & 63.4 & 63.4 & 36 & 36 & 149.5 \\ 117.1 & 114.8 & 99 & 99.2 & 57.9 & 57.9 & 35 & 35 & 114.8 \\ 118.7 & 116.4 & 99.2 & 99.7 & 58.2 & 58.2 & 35.4 & 35.4 & 116.4 \\ 65.1 & 63.4 & 57.9 & 58.2 & 46.8 & 46.7 & 37.1 & 37.1 & 63.4 \\ 65.1 & 63.4 & 57.9 & 58.2 & 46.7 & 46.8 & 37.1 & 37.1 & 63.4 \\ 36.8 & 36 & 35 & 35.4 & 37.1 & 37.1 & 39.3 & 39.3 & 36 \\ 36.8 & 36 & 35 & 35.4 & 37.1 & 37.1 & 39.3 & 39.3 & 36 \\ 152.3 & 149.5 & 114.8 & 116.4 & 63.4 & 63.4 & 36 & 36 & 149.5 \end{pmatrix},$$

$$S_{40} = \begin{pmatrix} 239.9 & 236.5 & 210.6 & 211.5 & 135.9 & 135.9 & 88 & 88 & 236.5 \\ 236.5 & 233.3 & 207.8 & 208.7 & 134.2 & 134.2 & 87 & 87 & 233.3 \\ 210.6 & 207.8 & 197.9 & 197.8 & 135 & 135 & 91.8 & 91.8 & 207.8 \\ 211.5 & 208.7 & 197.8 & 197.9 & 134.9 & 134.9 & 91.4 & 91.4 & 208.7 \\ 135.9 & 134.2 & 135 & 134.9 & 110.5 & 110.4 & 86.7 & 86.7 & 134.2 \\ 135.9 & 134.2 & 135 & 134.9 & 110.4 & 110.5 & 86.7 & 86.7 & 134.2 \\ 88 & 87 & 91.8 & 91.4 & 86.7 & 86.7 & 86.1 & 86 & 87 \\ 88 & 87 & 91.8 & 91.4 & 86.7 & 86.7 & 86 & 86.1 & 87 \\ 236.5 & 233.3 & 207.8 & 208.7 & 134.2 & 134.2 & 87 & 87 & 233.3 \end{pmatrix},$$

$$S_{80} = \begin{pmatrix} 318.2 & 315.6 & 288.6 & 289.8 & 195 & 195 & 137.7 & 137.7 & 315.5 \\ 315.6 & 313.2 & 286.5 & 287.7 & 194 & 194 & 137.3 & 137.3 & 313.1 \\ 288.6 & 286.5 & 272.4 & 273 & 189.3 & 189.3 & 137.5 & 137.5 & 286.4 \\ 289.8 & 287.7 & 273 & 273.1 & 189.9 & 189.9 & 137.8 & 137.8 & 287.6 \\ 195 & 194 & 189.3 & 189.9 & 148 & 147.9 & 117.9 & 117.9 & 193.9 \\ 195 & 194 & 189.3 & 189.9 & 147.9 & 148 & 117.9 & 117.9 & 193.9 \\ 137.7 & 137.3 & 137.5 & 137.8 & 117.9 & 117.9 & 140.9 & 140.9 & 137.1 \\ 137.7 & 137.3 & 137.5 & 137.8 & 117.9 & 117.9 & 140.9 & 140.9 & 137.1 \\ 315.5 & 313.1 & 286.4 & 287.6 & 193.9 & 193.9 & 137.1 & 137.1 & 313 \end{pmatrix},$$

and

$$S_{120} = \begin{pmatrix} 272 & 270.7 & 248.7 & 249.5 & 169.2 & 169.2 & 127.2 & 127.2 & 270.6 \\ 270.6 & 269.4 & 247.6 & 248.4 & 168.9 & 168.9 & 127.1 & 127.1 & 269.3 \\ 248.7 & 247.6 & 235.5 & 235.9 & 164.1 & 164.1 & 126.1 & 126.1 & 247.5 \\ 249.5 & 248.4 & 235.9 & 236.4 & 164.3 & 164.3 & 126.3 & 126.3 & 248.3 \\ 169.2 & 168.9 & 164.1 & 164.3 & 129.7 & 129.7 & 107.3 & 107.3 & 168.7 \\ 169.2 & 168.9 & 164.1 & 164.3 & 129.7 & 129.7 & 107.3 & 107.3 & 168.7 \\ 127.2 & 127.1 & 126.1 & 126.3 & 107.3 & 107.3 & 121.9 & 121.9 & 126.9 \\ 127.2 & 127.1 & 126.1 & 126.3 & 107.3 & 107.3 & 121.9 & 121.9 & 126.9 \\ 270.6 & 269.3 & 247.5 & 248.3 & 168.7 & 168.7 & 126.9 & 126.9 & 269.3 \end{pmatrix}.$$

We see that the variance for the Cash variable is much greater than for the other variables, for all t. This is expected since the Cash closing prices are more variable than the futures. This is clearly seen by looking at the correlation matrix $\boldsymbol{r}_t$. Here we provide the correlation matrices after ineteryention for $t = 20$ and $t = 80$.

$$r_{20} = \begin{pmatrix} 1 & 0.999 & 0.944 & 0.954 & 0.764 & 0.764 & 0.471 & 0.471 & 0.999 \\ 0.999 & 1 & 0.943 & 0.953 & 0.758 & 0.758 & 0.469 & 0.469 & 0.999 \\ 0.944 & 0.943 & 1 & 0.999 & 0.851 & 0.851 & 0.561 & 0.561 & 0.943 \\ 0.954 & 0.953 & 0.999 & 1 & 0.852 & 0.852 & 0.566 & 0.566 & 0.953 \\ 0.764 & 0.758 & 0.851 & 0.852 & 1 & 0.999 & 0.866 & 0.866 & 0.758 \\ 0.764 & 0.758 & 0.851 & 0.852 & 0.999 & 1 & 0.866 & 0.866 & 0.758 \\ 0.471 & 0.469 & 0.561 & 0.566 & 0.866 & 0.866 & 1 & 0.999 & 0.469 \\ 0.471 & 0.469 & 0.561 & 0.566 & 0.866 & 0.866 & 0.999 & 1 & 0.469 \\ 0.999 & 0.999 & 0.943 & 0.953 & 0.758 & 0.758 & 0.469 & 0.469 & 1 \end{pmatrix}$$

and

$$r_{80} = \begin{pmatrix} 1 & 0.999 & 0.980 & 0.982 & 0.898 & 0.898 & 0.651 & 0.651 & 0.999 \\ 0.999 & 1 & 0.981 & 0.982 & 0.901 & 0.901 & 0.653 & 0.653 & 0.999 \\ 0.980 & 0.981 & 1 & 0.999 & 0.943 & 0.943 & 0.702 & 0.702 & 0.981 \\ 0.982 & 0.982 & 0.999 & 1 & 0.943 & 0.943 & 0.702 & 0.702 & 0.982 \\ 0.898 & 0.901 & 0.943 & 0.943 & 1 & 0.999 & 0.816 & 0.816 & 0.901 \\ 0.898 & 0.901 & 0.943 & 0.943 & 0.999 & 1 & 0.816 & 0.816 & 0.901 \\ 0.651 & 0.653 & 0.702 & 0.702 & 0.816 & 0.816 & 1 & 0.999 & 0.653 \\ 0.651 & 0.653 & 0.702 & 0.702 & 0.816 & 0.816 & 0.999 & 1 & 0.653 \\ 0.999 & 0.999 & 0.981 & 0.982 & 0.901 & 0.901 & 0.653 & 0.653 & 1 \end{pmatrix}$$

We see that the correlation between Cash and 3 months is very high compared with the one of 15 months and 27 months. This is further illustrated in Figure 6.10.

Let $r_{13,t}, r_{35,t}, r_{57,t}$ denote the correlations of the variables Cash with 3 months, 3 months with 15 months, and 15 months with 27 months respec-

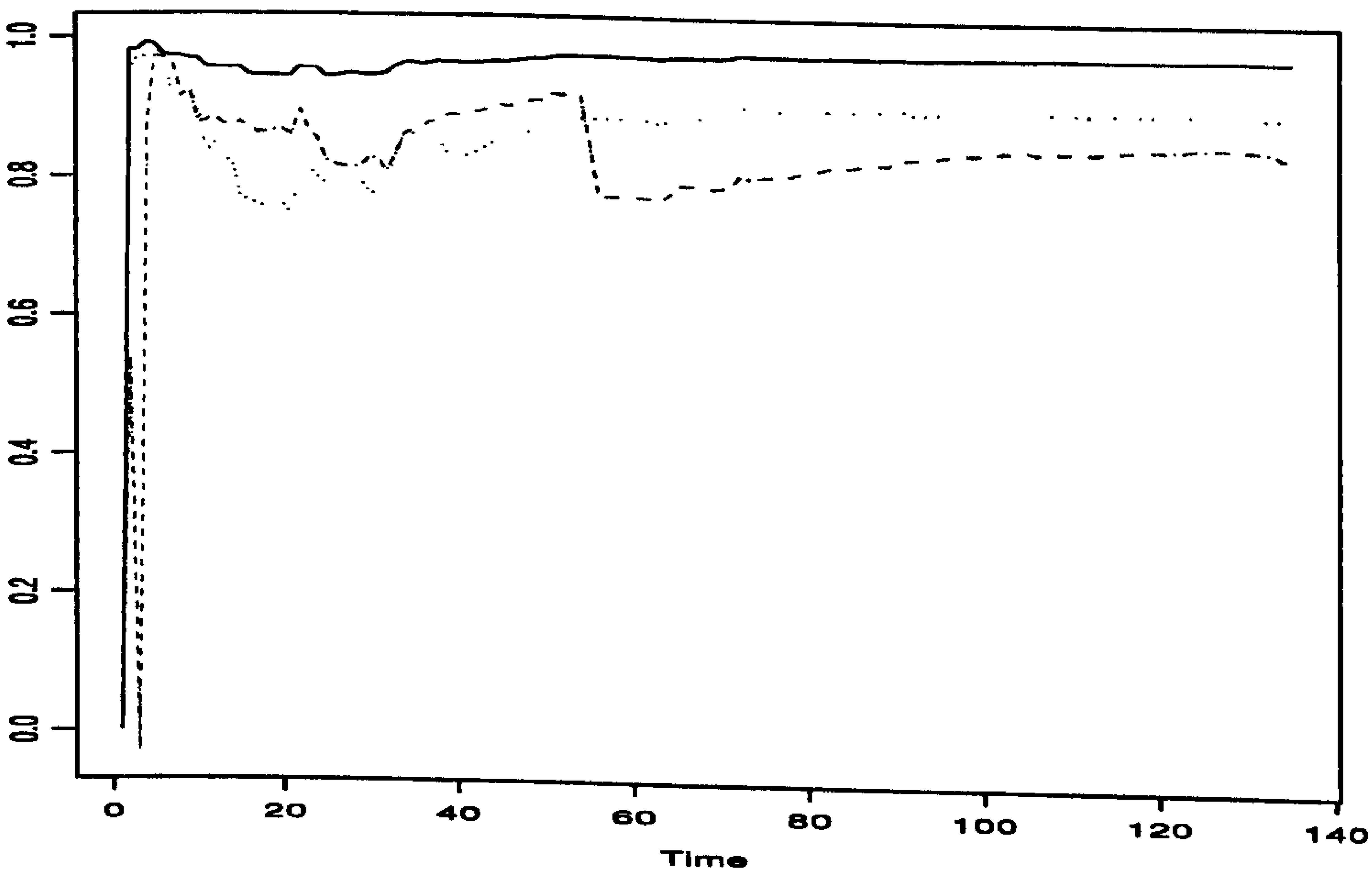

Figure 6.10: Comparison of the correlation of the variables

tively. Figure 6.10 draws these correlations for time $t = 1, \ldots, 135$. The three lines correspond to $r_{13,t}, r_{35,t}, r_{57,t}$ as read from the top.

We observe that after $t = 60$, $r_{57,t} < r_{35,t} < r_{13,t}$. The explanation of this is that the Cash and 3 months variables are more affected by the collected data, while the 15 months and 27 months variables are subject to more qualitative factors not available at each time t. This is shown in Figure 6.1 where the Cash and 3 months variables are smoother than the other two. The Cash is more variable and this is seen in the above variance matrices $\boldsymbol{S}_t$, correlation matrices $\boldsymbol{r}_t$, and the Figure 6.1. Indeed, this figure shows clearly that the range of Cash is wider than the respective of 27 months, with the first one being [1396, 1736], while the later [1493, 1568].

These agree perfectly with the discussed forms of information for capital markets in [11]. According to this, information may be divided into three forms. The first one is the *weak form*, in which the information set is just historical prices. The second one is the *semi-strong form*, which comprises additional qualitative information that is widely available (e.g. announcements of annual earnings). The third one is the *strong form*, which only provides monopolistic access for some investors or groups. In our example, the Cash variable comprises information in the weak form, the 3 months variable in the semi-strong form, and the 15, 27 months variables in the strong form. This indicates that forecasts for the last two variables that are based merely to previous data, may be very poor. However, in our model and for the specified time the forecasts of all variables after intervention have a similar performance. This is further illustrated to the following graphs.

Figure 6.11 shows the one step forecast after intervention of the variables Cash and 3 months (top graph) and Cash with 15 months (bottom graph). The continuous line shows the Cash forecasts, while the dotted line shows the other variables. Figure 6.12 shows the corresponding one step errors. Similar comments related with the description of this graph apply as in Figure 6.11. We see, from both figures, that most of the forecasts of the variables 3 months and 15 months are close to these for the variable Cash. This shows a good performance of the model. There is a significant number of forecasts for the variables 3 months and 15 months that the corresponding one step errors are smaller than they are for the variable Cash.

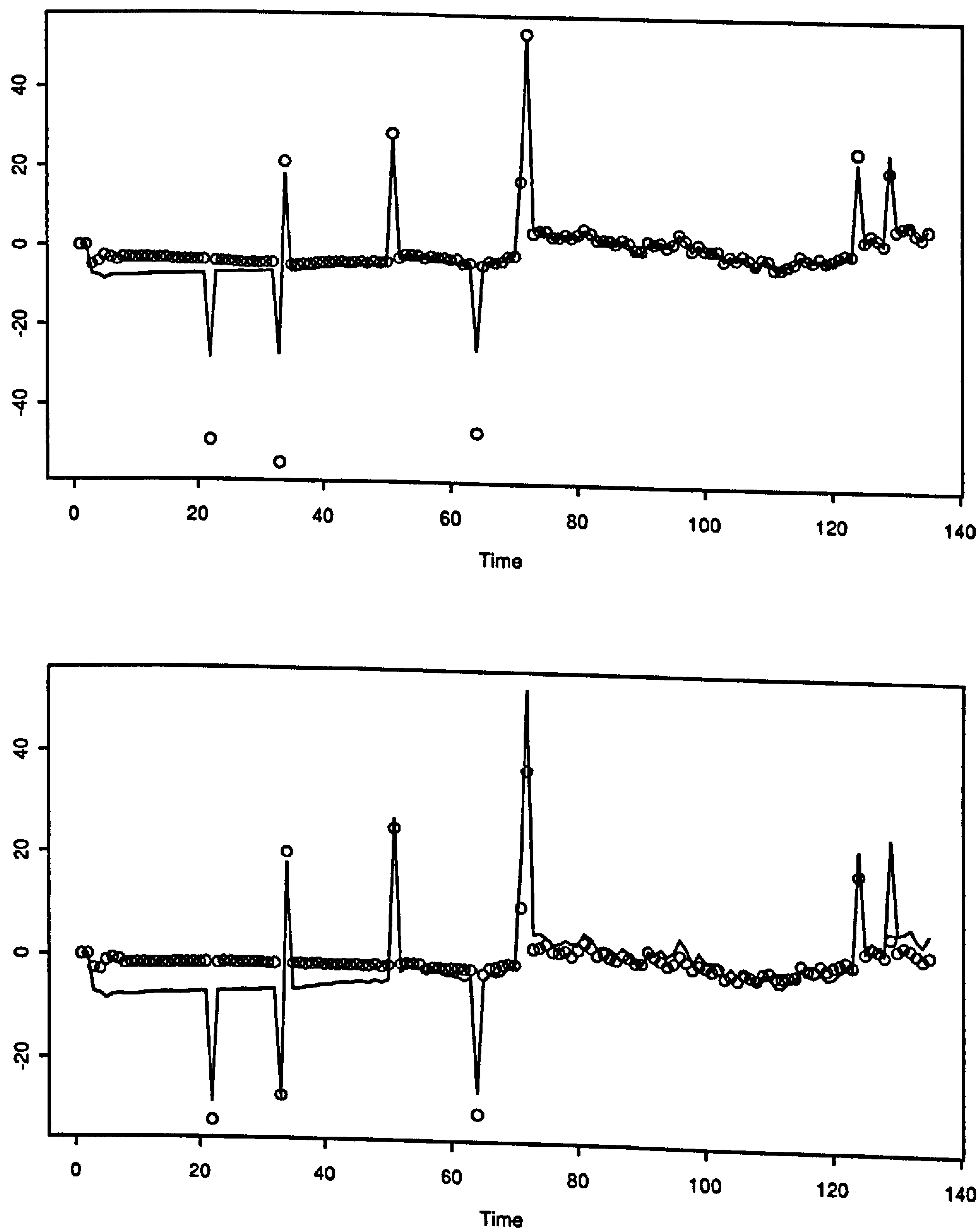

Figure 6.11: Comparison of one step forecasts between the variables

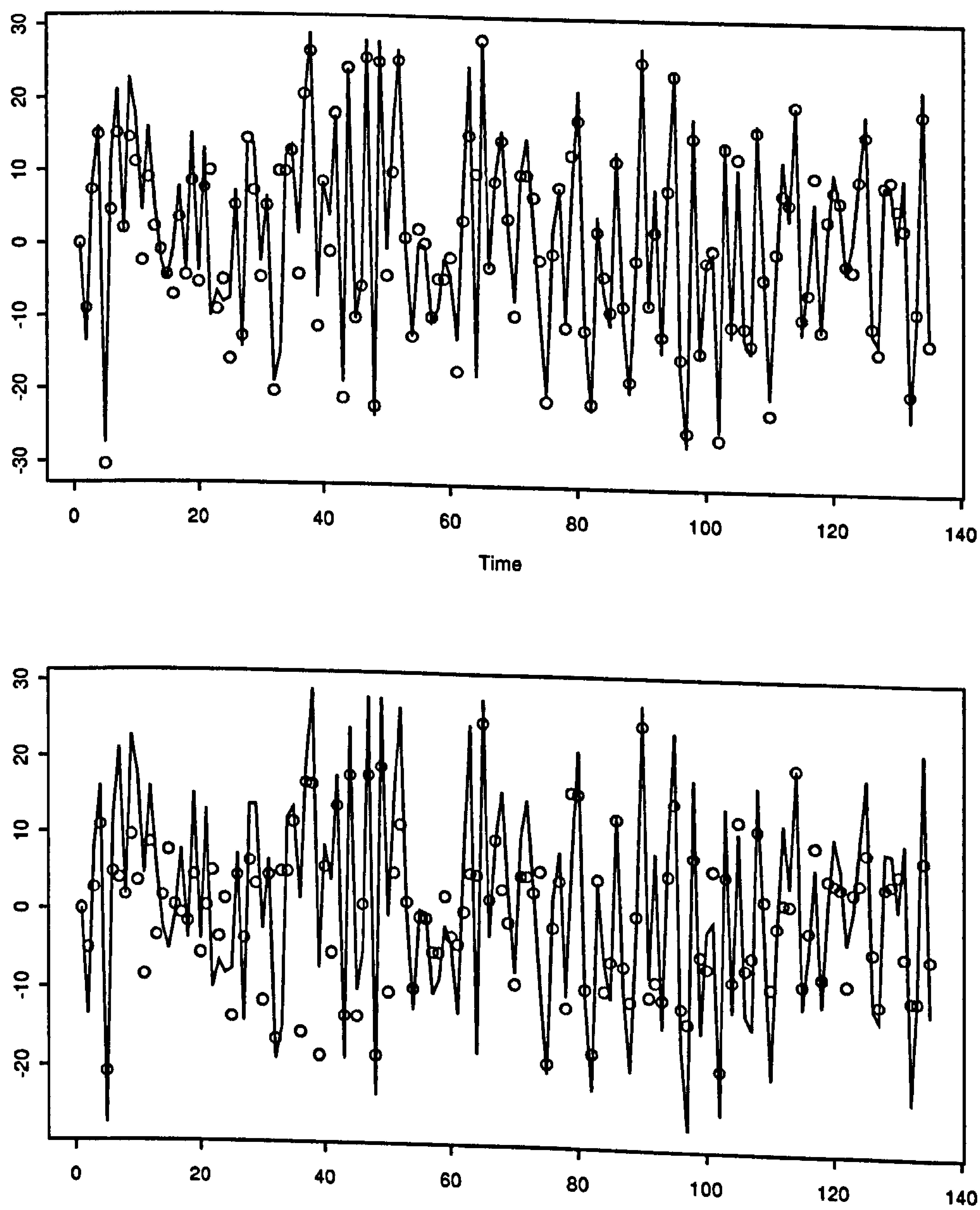

Figure 6.12: Comparison of one step errors between the variables

CHAPTER 7

Time Varying Variances

7.1 Introduction

In this thesis so far the unknown observational variance matrix, Σ, has been assumed constant over time. This chapter considers the problem of time varying variances. The problem is well known and is addressed in [51, chapters 10,16] and in [36, 37]. The observational variance may well change over time. The material of this chapter is to be found in [41]. For the first time deterministic variance laws are developed for the general multivariate DLM, in Section 7.2. Section 7.3 investigates the important problem of stochastic changes of the observational variance matrix. A generalization of the matrix beta distribution is introduced avoiding the use of a common discount factor, thus providing comprehensive variance evolution throughout the various elements of the observational variance matrix. This follows [41]. Section 7.4 discusses some practical aspects of the problem and in the last section an illustration is presented.

7.2 Deterministic Variance Laws

A deterministic variance law is a term referring to the dependence of time in the variance by a deterministic function. That is, the variance Σ_t does not include stochastic changes over time. This is a very practical case.

In this thesis the general multivariate DLM as developed in Chapter 5, is illustrated. The reason for this is (a) this is a more general model than the currently used CCM, and more widely applicable, (b) it allows analytical recurrence relationships under the weak assumptions, therefore is preferable to approximations like [4], and (c) it allows a very flexible analysis for incorporating time varying observational variances. (c) is explored in this section and it is shown why matrix DLMs like the CCM and ECCM do not allow for a similar analysis. Also other possible approaches are investigated.

7.2.1 The Linear Variance Law

Consider the model

$$\boldsymbol{Y}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim [\mathbf{0}, \Sigma_t], \tag{7.1}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim [\mathbf{0}, \boldsymbol{W}_t], \tag{7.1$'$}$$

where the quantities $\boldsymbol{Y}_t$, $\boldsymbol{F}_t$, $\boldsymbol{\theta}_t$, $\boldsymbol{G}_t$, $\boldsymbol{W}_t$, $\boldsymbol{\nu}_t$, $\boldsymbol{\omega}_t$ are as defined in Section 5.3 and the variance matrix Σ_t is factorised as

$$\Sigma_t = \boldsymbol{L}_t\Sigma\boldsymbol{L}_t', \tag{7.2}$$

where $\boldsymbol{L}_t$ is a known $r \times r$ non-singular matrix.

So Σ_t is known up to an unknown constant matrix Σ. Model (7.1), (7.1′) can be transformed to another one with an unknown constant variance matrix Σ. By defining

$$\boldsymbol{Y}_t^* = \boldsymbol{L}_t^{-1}\boldsymbol{Y}_t, \qquad \boldsymbol{F}_t^* = \boldsymbol{F}_t\boldsymbol{L}_t^{\prime -1}, \qquad \boldsymbol{\nu}_t^* = \boldsymbol{L}_t^{-1}\boldsymbol{\nu}_t,$$

we have the resulting model

$$\boldsymbol{Y}_t^* = \boldsymbol{F}_t^{*\prime}\boldsymbol{\theta}_t + \boldsymbol{\nu}_t^*, \qquad \boldsymbol{\nu}_t^* \sim [0, \Sigma], \tag{7.3}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim [0, \boldsymbol{W}_t]. \tag{7.3′}$$

Note that the evolution equations, (7.1′) and (7.3′), of the two models are identical, therefore inferences about the state $\boldsymbol{\theta}_t$ can be directly derived by model (7.3), (7.3′). This class of models is explicitly developed in Chapter 5.

Denote the expectation of Σ given D_t by $\boldsymbol{S}_t^*$ and the expectation of Σ_t given D_t by $\boldsymbol{S}_t$.

Assume as usual that the posterior at time $t-1$ is

$$(\boldsymbol{\theta}_{t-1}|D_{t-1}) \sim [\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}],$$

for some known quantities $\boldsymbol{m}_{t-1}$, $\boldsymbol{C}_{t-1}$.

Using simple mean and variance calculus we have that the prior at t is given by

$$(\boldsymbol{\theta}_t|D_{t-1}) \sim [\boldsymbol{a}_t, \boldsymbol{R}_t],$$

where $\boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1}$ and $\boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t' + \boldsymbol{W}_t$.

Similarly the one-step forecast at t is

$$(\boldsymbol{Y}_t^*|D_{t-1}) \sim [\boldsymbol{f}_t^*, \boldsymbol{Q}_t^*],$$

where

$$f_t^* = F_t^{*\prime} a_t = L_t^{-1} f_t$$
$$Q_t^* = F_t^{*\prime} R_t F_t^* + S_{t-1}^* = L_t^{-1} Q_t L_t^{\prime -1},$$

and $f_t = F_t' a_t$, $Q_t = F_t' R_t F_t + S_{t-1}$.

The adaptive factor A_t^* is

$$\begin{aligned} A_t^* &= R_t F_t^* Q_t^{*-1} \\ &= R_t F_t L_t^{\prime -1} L_t' Q_t^{-1} L_t \\ &= R_t F_t Q_t^{-1} L_t \\ &= A_t L_t. \end{aligned}$$

Also

$$e_t^* = Y_t^* - f_t^* = L_t^{-1}(Y_t - f_t) = L_t^{-1} e_t.$$

So $A_t^* e_t^* = A_t e_t$,

$$m_t = a_t + A_t^* e_t^* = a_t + A_t e_t$$

and

$$\begin{aligned} C_t &= R_t - A_t^* Q_t^* A_t^{*\prime} \\ &= R_t - A_t L_t L_t^{-1} Q_t L_t^{\prime -1} L_t' A_t' \\ &= R_t - A_t Q_t A_t'. \end{aligned}$$

Therefore, the posterior at t is

$$(\theta_t | D_t) \sim [m_t, C_t],$$

with m_t, C_t as calculated above.

The recurrence of S_t follows from Section 6.3.3

$$S_t^* = S_{t-1}^* + N_t^{-1/2} S_{t-1}^{*1/2} (Q_t^{*-1/2} U_t e_t^* e_t^{*\prime} U_t Q_t^{*-1/2} - U_t) S_{t-1}^{*1/2} N_t^{-1/2}$$

and from

$$S_t = L_t S_t^* L_t',$$

where $N_t = N_{t-1} + U_t$.

So

$$\begin{aligned} S_t &= L_t L_{t-1}^{-1} S_{t-1} L_{t-1}^{\prime -1} L_t' + L_t N_t^{-1/2} (L_{t-1}^{-1} S_{t-1} L_{t-1}^{\prime -1})^{1/2} \\ &\quad \times [(L_t^{-1} Q_t L_t^{\prime -1})^{-1/2} U_t L_t^{-1} e_t e_t' L_t^{\prime -1} U_t (L_t^{-1} Q_t L_t^{\prime -1})^{-1/2} - U_t] \\ &\quad \times (L_{t-1}^{-1} S_{t-1} L_{t-1}^{\prime -1})^{1/2} N_t^{-1/2} L_t'. \end{aligned}$$

The assumptions used for the above analysis are those of Sections 5.4 and 6.3.3, namely

$$\theta_t - A_t Y_t \perp_2 L_t^{-1} Y_t | \Sigma, D_{t-1}, \tag{7.4}$$

$$\mathrm{vech}(\Sigma - \mathcal{A}_t L_t^{-1} e_t e_t' L_t^{\prime -1} \mathcal{A}_t') \perp_1 L_t^{-1} Y_t | D_{t-1}, \tag{7.5}$$

where $\mathcal{A}_t = N_t^{-1/2} (L_{t-1}^{-1} S_{t-1} L_{t-1}^{\prime -1})^{1/2} (L_t^{-1} Q_t L_t^{\prime -1})^{-1/2}$.

Summarizing we have the following theorem.

Theorem 7.1. *In the multivariate DLM (7.1), (7.1′) using assumptions (7.4), (7.5), one-step forecast and posterior distributions are partially given, for each t, as follows:*

(a) Posterior at $t-1$:

For some mean m_{t-1} and variance matrix C_{t-1},

$$(\theta_{t-1} | D_{t-1}) \sim [m_{t-1}, C_{t-1}].$$

(b) Prior at t:

$$(\boldsymbol{\theta}_t|D_{t-1}) \sim [\boldsymbol{a}_t, \boldsymbol{R}_t],$$

where

$$\boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1} \qquad \textit{and} \qquad \boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t' + \boldsymbol{W}_t.$$

(c) One-step forecast:

$$(\boldsymbol{Y}_t|D_{t-1}) \sim [\boldsymbol{f}_t, \boldsymbol{Q}_t],$$

where

$$\boldsymbol{f}_t = \boldsymbol{F}_t'\boldsymbol{a}_t \qquad \textit{and} \qquad \boldsymbol{Q}_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + \boldsymbol{S}_{t-1}.$$

(d) Posterior at t:

$$(\boldsymbol{\theta}_t|D_t) \sim [\boldsymbol{m}_t, \boldsymbol{C}_t],$$

with

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t\boldsymbol{e}_t \qquad \textit{and} \qquad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{Q}_t\boldsymbol{A}_t',$$

where

$$\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{Q}_t^{-1} \qquad \textit{and} \qquad \boldsymbol{e}_t = \boldsymbol{Y}_t - \boldsymbol{f}_t.$$

(e) Updating of the variance estimate:

$$\begin{aligned} \boldsymbol{S}_t &= \boldsymbol{L}_t\boldsymbol{L}_{t-1}^{-1}\boldsymbol{S}_{t-1}\boldsymbol{L}_{t-1}'^{-1}\boldsymbol{L}_t' + \boldsymbol{L}_t\boldsymbol{N}_t^{-1/2}(\boldsymbol{L}_{t-1}^{-1}\boldsymbol{S}_{t-1}\boldsymbol{L}_{t-1}'^{-1})^{1/2} \\ &\quad \times[(\boldsymbol{L}_t^{-1}\boldsymbol{Q}_t\boldsymbol{L}_t'^{-1})^{-1/2}\boldsymbol{U}_t\boldsymbol{L}_t^{-1}\boldsymbol{e}_t\boldsymbol{e}_t'\boldsymbol{L}_t'^{-1}\boldsymbol{U}_t(\boldsymbol{L}_t^{-1}\boldsymbol{Q}_t\boldsymbol{L}_t'^{-1})^{-1/2} - \boldsymbol{U}_t] \\ &\quad \times(\boldsymbol{L}_{t-1}^{-1}\boldsymbol{S}_{t-1}\boldsymbol{L}_{t-1}'^{-1})^{1/2}\boldsymbol{N}_t^{-1/2}\boldsymbol{L}_t' \end{aligned}$$

and

$$\boldsymbol{N}_t = \boldsymbol{N}_{t-1} + \boldsymbol{U}_t.$$

Now we show that such a methodology is not applicable to the matrix normal DLMs. Consider the ECCM, as it was developed in Section 4.4, but now Σ is replaced by a time varying Σ_t as in equation (7.2). The model is given by

$$Y_t' = F_t'\Theta_t + \nu_t', \qquad \nu_t' \sim \mathrm{N}[0, V_t, \Sigma_t],$$

$$\Theta_t = G_t\Theta_{t-1} + \Omega_t, \qquad \Omega_t \sim \mathrm{N}[\mathbf{O}, W_t, \Sigma_t],$$

and

$$(\Theta_0, \Sigma_0 | D_0) \sim \mathrm{NGW}^{-1}[m_0, C_0, S_0, N_0, m_0].$$

In order to obtain another matrix model such that the observational right variance matrix is constant, we have to apply the transformation

$$Y_t^* = L_t'^{-1}Y_t, \qquad \Theta_t^* = \Theta_t L_t^{-1}, \qquad \nu_t^* = L_t'^{-1}\nu_t, \qquad \Omega_t^* = \Omega_t L_t^{-1}.$$

The observation equation becomes

$$Y_t^{*\prime} = F_t'\Theta_t^* + \nu_t^{*\prime}, \qquad \nu_t^{*\prime} \sim \mathrm{N}[0, V_t, \Sigma].$$

However, a respective evolution equation cannot be derived since

$$\Theta_t L_t^{-1} = G_t\Theta_{t-1}L_t^{-1} + \Omega_t L_t^{-1},$$

implies

$$\Theta_t^* = G_t\Theta_{t-1}^* L_{t-1}L_t^{-1} + \Omega_t^*.$$

It follows that only when $L_t = L$ is it possible to work it out, but in this case Σ_t is constant over time and there is nothing gained by the analysis.

Returning to model (7.1), (7.1′), we discuss the specification of L_t. A practical approach is to introduce a positive real valued sequence $\{\delta_t\}$ such that

$$L_t = \sqrt{\delta_t}L_{t-1}. \tag{7.6}$$

The motivation of this is as follows. Starting with $\boldsymbol{L}_0 = \boldsymbol{I}$ we construct the sequence $\{\boldsymbol{L}_t\}$, using equation (7.6). The factor δ_t measures the dispersion of $\boldsymbol{L}_t$ from $\boldsymbol{L}_{t-1}$. If $0 < \delta_t < 1$, $\|\boldsymbol{L}_t\| < \|\boldsymbol{L}_{t-1}\|$, while if $\delta_t > 1$, $\|\boldsymbol{L}_t\| > \|\boldsymbol{L}_{t-1}\|$, for any matrix norm $\|\cdot\|$, (see Appendix A.5). If $\delta_t = 1$, $\boldsymbol{L}_t = \boldsymbol{L}_{t-1}$, and there is no change in the variance from time $t-1$ to t. This is very practical since the variance $\boldsymbol{\Sigma}_t$ is not likely to change at every consecutive point of time. The sequence $\{\delta_t\}$ can be defined as

$$\delta_t = \delta_{t-1} + \epsilon_t,$$

where ϵ_t is a quantity that has to be specified by the modeller for every t, according to relevant prior knowledge or expectation. A more advanced approach is to assume that $\epsilon_t \sim [\varepsilon_t, V_{\epsilon,t}]$ for some known moments ε_t and $V_{\epsilon,t}$. However, this approach is not adopted here.

Another extension to the above methodology is to consider a sequence of diagonal matrices, $\{\boldsymbol{\Delta}_t\}$, instead of $\{\delta_t\}$. The above analysis, using $\{\delta_t\}$ as in (7.6), assumes that the elements of $\boldsymbol{L}_t$, thus the elements of $\boldsymbol{\Sigma}_t$, change over time with the same rate. This is not always desirable, that is why $\boldsymbol{\Delta}_t$ is to be introduced. Now equation (7.6) is replaced by

$$\boldsymbol{L}_t = \boldsymbol{\Delta}_t^{1/2} \boldsymbol{L}_{t-1}$$

and

$$\boldsymbol{\Delta}_t = \boldsymbol{\Delta}_{t-1} + \boldsymbol{E}_t,$$

where $\boldsymbol{\Delta}_t = \text{diag}\{\delta_{1t}, \dots, \delta_{rt}\}$ and $\boldsymbol{E}_t = \text{diag}\{\epsilon_{1t}, \dots, \epsilon_{rt}\}$, for some known positive quantities δ_{it}, ϵ_{it}, $(i = 1, \dots, r)$. A stochastic evolution of $\boldsymbol{\Delta}_t$ may be considered, but again this is not developed here.

7.2.2 General Deterministic Variance Laws

In Section 7.2.1 a time dependent variance $\Sigma_t = L_t \Sigma L_t'$ was considered and DLM analysis was possible within the framework of Chapter 5. The question is: "Will it always be possible to find a suitable matrix L_t for any variance Σ_t such that $\Sigma_t = L_t \Sigma L_t'$?" Or more mathematically, "given an unknown variance matrix Σ_t does there exist a non-singular matrix L_t such that $\Sigma_t = L_t \Sigma L_t'$?" And if so, "will it be possible using the approach of Section 7.2.1 to calculate this L_t?" In theory the answer to these questions is negative. To see this, consider a general unknown $\Sigma_t = \{\sigma_{ij,t}\}$, an unknown constant $\Sigma = \{\sigma_{ij}\}$ and a non-singular matrix $L_t = \{l_{ij,t}\}$ that does not include any random quantities, $(i, j = 1, \dots, r)$. By equating $\Sigma_t = L_t \Sigma L_t'$ we construct a system of $r(r+1)/2$ non-linear equations with r^2 unknown quantities.

$$\sum_{k=1}^{r} \sum_{m=1}^{r} l_{im,t} l_{jk,t} \sigma_{mk} = \sigma_{ij,t},$$

for $1 \leq i, j \leq r$. The existence of at least one solution of $l_{ij,t}$ of the above system ensures that the construction $\Sigma_t = L_t \Sigma L_t'$ is always possible. Unfortunately this is not possible, since in most of the cases (for most of the unknown Σ_t) the elements of L_t will depend upon either or both the elements of Σ_t, Σ. Thus, in general the above system will provide L_t as random quantities. This is not to say that the development of Section 7.2.1 is unimportant. In practice we are not interested in capturing the exact variance Σ_t, rather it is to track the impact that it has on the current observations and to build a more realistic and flexible model. This is done quite well within the framework of Section 7.2.1. Next, we attempt a more general analysis that leads to a general estimate of Σ_t.

Consider the model (7.1), (7.1′), where Σ_t is any variance matrix. Suppose that the system matrix G_t and the system variance W_t satisfy

$$G_t = \text{block diag}\{G_{0t}, G_{1t}, \dots, G_{kt}\},$$

$$W_t = \text{block diag}\{W_{0t}, W_{1t}, \dots, W_{kt}\},$$

for some known matrices G_{it}, W_{it}, $(i = 0, 1, \dots, k)$ and $k \geq 0$.

Let $\mathcal{V}$ be the space of all $r \times r$ variance matrices and $\mathcal{S} = (a, b) \supset [0, s]$, for $s > 1$. Consider now the continuous matrix valued function $F : \mathcal{S} \longrightarrow \mathcal{V}$ such that $F(t) = \Sigma_t$, for every $t \in \mathbb{N}^*_{[s+1]} = \{1, \dots, [s]\}$, where $[x]$ denotes the integral part of the real number x. Assuming that there exist all $[\partial^i F/\partial t^i]_{t=t_0}$, $i = 1, 2, \dots$, for $t_0 \in \mathcal{S}$ we can use the Taylor expansion as

$$F(t) = \sum_{i=0}^{\infty} \frac{(t - t_0)^i}{i!} \left[\frac{\partial^i F(t)}{\partial t^i}\right]_{t=t_0},$$

where as usual it is set $[\partial^0 F/\partial t^0]_{t=t_0} = F(t_0)$.

So taking $t \in \mathbb{N}$,

$$\Sigma_t \approx \sum_{i=0}^{k} \frac{(t - t_0)^i}{i!} \Sigma_0{}^{(i)},$$

where $k \in \mathbb{N}$ and $\Sigma_0{}^{(i)} = [\partial^i F/\partial t^i]_{t=t_0}$ has been calculated before considering t as integer, so that the partial derivatives are sensible. In practice it will suffice to consider $k = 1$, or $k = 2$ for 1st/2nd order Taylor approximation.

Assuming $k \leq r - 1$ we obtain the following decomposition of the model (7.1), (7.1′)

$$Y_{it} = F'_{it}\theta_{it} + \nu_{it}, \qquad \nu_{it} \sim \left[0, \frac{(t - t_0)^i}{i!} \Sigma_0{}^{(i)}\right], \tag{7.7}$$

$$\theta_{it} = G_{it}\theta_{i,t-1} + \omega_{it}, \qquad \omega_{it} \sim [0, W_{it}], \tag{7.7′}$$

where

$$Y_t = \sum_{i=0}^{k} Y_{it}, \qquad F'_t = (F'_{0t}, F'_{1t}, \dots, F'_{kt}), \qquad \nu_t = \sum_{i=0}^{k} \nu_{it},$$

$$\theta'_t = (\theta'_{0t}, \theta'_{1t}, \dots, \theta'_{kt}), \qquad \omega'_t = (\omega'_{0t}, \omega'_{1t}, \dots, \omega'_{kt}),$$

and $\dim(Y_{it}) = r \times 1$, $\dim(F_{it}) = n_i \times r$, $\dim(\nu_{it}) = r \times 1$, $\dim(\theta_{it}) = n_i \times 1$, $\dim(\omega_{it}) = n_i \times 1$, $\sum_{i=0}^{k} n_i = n$, $(i = 0, 1, \dots, k)$.

For more details on model decomposition see Section 2.5 and [51, chapter 6].

We also employ the respective assumptions for each of the $k+1$ decomposed DLMs, as it was done for the individual model (7.1), (7.1), see also equations (7.4), (7.5).

Denote with S_t the estimate of Σ_t given D_t, and with $S_{t,0}^{(i)}$ the estimate of $\Sigma_0^{(i)}$ given $D_{i,t} = \{D_{i,t-1}, Y_{it}\}$. So

$$S_t \approx \sum_{i=0}^{k} \frac{(t-t_0)^i}{i!} S_{t,0}^{(i)}. \tag{7.8}$$

The final step is to calculate the estimates $S_{t,0}^{(i)}$ for $0 \le i \le k$. Each one of the $k+1$ models are in the model class of Section 7.2.1. This is true because each one of the unknown variances $\Sigma_0^{(i)}$ is constant over time. So referring to the previous section, we define the time varying variance matrix for the ith model

$$\Sigma_{t,0}^{(i)} = \frac{(t-t_0)^i}{i!} \Sigma_0^{(i)}$$

and so

$$L_{t,i} = \left(\frac{|t-t_0|^i}{i!} \right)^{1/2} I,$$

with $\Sigma_{t,0}^{(i)} = L_{t,i} \Sigma_0^{(i)} L'_{t,i}$, for $i = 0, 1, \dots, k$. Assuming the case of no partial missing observations and setting $N_t = \text{diag}\{n_t, \dots, n_t\}$, we have from

Theorem 7.1

$$n_t S^{(i)}_{t,0} = \left| \frac{t-t_0}{t-t_0-1} \right|^i \left[n_{t-1} S^{(i)}_{t-1,0} + (S^{(i)}_{t-1,0})^{1/2} Q^{-1/2}_{i,t} e_{i,t} e'_{i,t} Q^{-1/2}_{i,t} (S^{(i)}_{t-1,0})^{1/2} \right], \tag{7.9}$$

where $Q_{i,t} = F'_{it} R_{i,t} F_{it} + S^{(i)}_{t-1,0}$, $R_{i,t} = G_{it} C_{i,t-1} G'_{it} + W_{it}$, and $C_{i,t-1}$ is the posterior at $t-1$ of $\theta_{i,t-1}$ given $D_{i,t-1}$, $(i = 0, 1, \ldots, k)$.

Summarizing all the above we have the following theorem.

Theorem 7.2. *Consider the multivariate DLM (7.1), (7.1′) with defining quantities*

$$G_t = block\ diag\{G_{0t}, G_{1t}, \ldots, G_{kt}\},$$
$$W_t = block\ diag\{W_{0t}, W_{1t}, \ldots, W_{kt}\},$$

for some known quantities G_{it}, W_{it} $(i = 0, 1, \ldots, k)$. Decomposing the above DLM into $k+1$ $(k \leq r-1)$ DLMs and employing assumptions (7.4), (7.5) for each one of the $k+1$ DLMs, one-step forecast and posterior distributions are partially given, for each t, as follows:

(a) Posterior at $t-1$:

For some mean m_{t-1} and variance matrix C_{t-1},

$$(\theta_{t-1} | D_{t-1}) \sim [m_{t-1}, C_{t-1}].$$

(b) Prior at t:

$$(\theta_t | D_{t-1}) \sim [a_t, R_t],$$

where

$$a_t = G_t m_{t-1} \qquad \text{and} \qquad R_t = G_t C_{t-1} G'_t + W_t.$$

(c) One-step forecast:

$$(\boldsymbol{Y}_t|D_{t-1}) \sim [\boldsymbol{f}_t, \boldsymbol{Q}_t],$$

where

$$\boldsymbol{f}_t = \boldsymbol{F}_t'\boldsymbol{a}_t \qquad \textit{and} \qquad \boldsymbol{Q}_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + \boldsymbol{S}_{t-1}.$$

(d) Posterior at t:

$$(\boldsymbol{\theta}_t|D_t) \sim [\boldsymbol{m}_t, \boldsymbol{C}_t],$$

with

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t\boldsymbol{e}_t \qquad \textit{and} \qquad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{Q}_t\boldsymbol{A}_t',$$

where

$$\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{Q}_t^{-1} \qquad \textit{and} \qquad \boldsymbol{e}_t = \boldsymbol{Y}_t - \boldsymbol{f}_t.$$

(e) Updating of the variance estimate:

$$\boldsymbol{S}_t \approx \sum_{i=0}^{k} \frac{(t-t_0)^i}{i!} \boldsymbol{S}_{t,0}^{(i)},$$

where

$$\begin{aligned} n_t\boldsymbol{S}_{t,0}^{(i)} &= \left|\frac{t-t_0}{t-t_0-1}\right|^i \Big[n_{t-1}\boldsymbol{S}_{t-1,0}^{(i)} + (\boldsymbol{S}_{t-1,0}^{(i)})^{1/2}\boldsymbol{Q}_{i,t}^{-1/2}\boldsymbol{e}_{i,t}\boldsymbol{e}_{i,t}' \\ &\quad \times \boldsymbol{Q}_{i,t}^{-1/2}(\boldsymbol{S}_{t-1,0}^{(i)})^{1/2}\Big], \end{aligned}$$

and

$$n_t = n_{t-1} + 1,$$

for any $t_0 > 0$, such that $|t - t_0| < \epsilon$, $\forall \epsilon > 0$.

Note that the estimate of $\boldsymbol{\Sigma}_t$ is only locally appropriate in an open neighborhood of t_0. There must be a reasonable evolution of t_0 for every t, in order to keep $|t - t_0| < \epsilon$, for all $\epsilon > 0$.

7.3 Stochastic Changes in Variance

Suppose that the variance Σ_t changes stochastically with time. In the univariate case variance discounting ideas as introduced by Ameen and Harrison [2], include the use of an inverse gamma/beta distribution for generating $\Sigma_t = \Sigma_t$ from Σ_{t-1}. The prior of Σ_t is an inverse gamma distribution with increased variance and the posterior of Σ_t is an inverse gamma with discounted degrees of freedom, $n_t = \beta n_{t-1} + 1$, where β is a discount factor. More details about the method are to be found in the above reference, in [51, chapter 10], and in [21, 22]. The method is fully considered and explored in its most general form in Section 7.3.1.

The above methodology was extended in [36, 37] based on matrix-beta distributions and applied in [35]. Also Uligh [46, 47] improved the method by introducing the singular-beta distribution and applied this to a stochastic volatility model. The important aspect of this is that even if the matrices are singular the conjugacy between the Wishart and the beta distributions still holds. However, the methods are susceptible due to scalar degrees of freedom in Wishart and beta distributions. So an implicit assumption of these methods is that no partial missing observations exist. The degrees of freedom are discounted at the same rate which is not always appropriate. In other words all the observations Y_{it}, ($\boldsymbol{Y}_t = (Y_{1t}, \ldots, Y_{rt})'$; $i = 1, \ldots, r$), have the same effect in the change of Σ_t. This assumption makes the Common Components Model too restrictive for general application. In the following section we extend the matrix beta distribution to incorporate a matrix of degrees of freedom, as in Chapter 4 for the Wishart distribution, and develop the theory of stochastic changes in variance using these distributions. We also consider the general multivariate DLM of Chapter 5 with time varying observational variances. Both methods are to be found in [41].

7.3.1 The Extended Common Components Model

First the matrix generalized beta distribution is introduced.

Suppose that $\boldsymbol{P}$ is an $r \times r$ symmetric positive definite matrix and a, b positive scalars. Then, the matrix-variate type I or matrix-variate (for short) non-singular beta distribution is defined by the density

$$p(\boldsymbol{P}) = c_1 |\boldsymbol{P}|^{a-(r+1)/2} |\boldsymbol{I} - \boldsymbol{P}|^{b-(r+1)/2}, \tag{7.10}$$

where $\mathbf{O} < \boldsymbol{P} < \boldsymbol{I}$ (both matrices $\boldsymbol{P}$, $\boldsymbol{I} - \boldsymbol{P}$ are positive definite) and $a, b > (r-1)/2$. The notation is $\boldsymbol{P} \sim \text{Beta}_r[a, b]$.

The proportionality constant, c_1, is given by

$$c_1 = [\beta_r(a,b)]^{-1} = \frac{\Gamma_r(a+b)}{\Gamma_r(a)\Gamma_r(b)},$$

where $\beta_r(\cdot)$ is the multivariate beta function and $\Gamma_r(\cdot)$ the multivariate gamma function (see Appendix B.3).

Lemma 7.1. *Let $\boldsymbol{P}$ be an $r \times r$ symmetric positive definite matrix and $\boldsymbol{N}_1$, $\boldsymbol{N}_2$ any $r \times r$ diagonal matrices with positive elements in their diagonal. Then, the function*

$$p(\boldsymbol{P}) = c|\boldsymbol{P}|^{\frac{trace(\boldsymbol{N}_1)}{2r} - 1} |\boldsymbol{I} - \boldsymbol{P}|^{\left(\frac{trace(\boldsymbol{N}_2)}{r} - r - 1\right)/2}, \tag{7.11}$$

where c is a constant not involving $\boldsymbol{P}$, defines a density, hence a distribution.

Proof. The proof is trivial, just by setting

$$a = \frac{1}{2}\left(r + \frac{\text{trace}(\boldsymbol{N}_1)}{r} - 1\right) \quad \text{and} \quad b = \frac{\text{trace}(\boldsymbol{N}_2)}{2r},$$

and using equation (7.10). □

The proportionality constant, c, is given by

$$c = \frac{\Gamma_r\left[\frac{1}{2}\left(r + \frac{\mathrm{trace}(\boldsymbol{N}_1+\boldsymbol{N}_2)}{r} - 1\right)\right]}{\Gamma_r\left[\frac{1}{2}\left(r + \frac{\mathrm{trace}(\boldsymbol{N}_1)}{r} - 1\right)\right]\Gamma_r\left[\frac{\mathrm{trace}(\boldsymbol{N}_2)}{2r}\right]}.$$

Definition 7.1. *Let $\boldsymbol{P}$ be an $r \times r$ symmetric positive definite matrix and $\boldsymbol{N}_1$, $\boldsymbol{N}_2$ diagonal $r \times r$ matrices with positive elements in their diagonals. $\boldsymbol{P}$ is said to follow the* **generalized beta** *distribution with $\boldsymbol{N}_1$ and $\boldsymbol{N}_2$ matrices of degrees of freedom if its density is given by equation (7.11). This may be written as $\boldsymbol{P} \sim GB\left[\frac{1}{2}\boldsymbol{N}_1, \frac{1}{2}\boldsymbol{N}_2\right]$.*

Now consider the ECCM with a time varying variance matrix $\boldsymbol{\Sigma}_t$ as defined in Chapter 4 by

$$\boldsymbol{Y}_t' = \boldsymbol{F}_t'\Theta_t + \boldsymbol{\nu}_t', \qquad \boldsymbol{\nu}_t' \sim \mathrm{N}[\boldsymbol{0}, V_t, \Sigma_t], \tag{7.12}$$

$$\Theta_t = \boldsymbol{G}_t\Theta_{t-1} + \boldsymbol{\Omega}_t, \qquad \boldsymbol{\Omega}_t \sim \mathrm{N}[\mathbf{O}, \boldsymbol{W}_t, \Sigma_t], \tag{7.12'}$$

and

$$(\Theta_0, \Sigma_0 | D_0) \sim \mathrm{NGW}^{-1}[\boldsymbol{m}_0, \boldsymbol{C}_0, \boldsymbol{S}_0, \boldsymbol{N}_0, m_0], \tag{7.13}$$

for some known defining parameters $\boldsymbol{m}_0$, $\boldsymbol{C}_0$, $\boldsymbol{S}_0$, $\boldsymbol{N}_0 = \mathrm{diag}\{n_{10}, \ldots, n_{r0}\}$.

Before we proceed we will need the following theorem, as appeared in [29, chapter 3].

Theorem 7.3. *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be $r \times r$ independent SPD random matrices, with $\boldsymbol{X} \sim W_r[n_1, \boldsymbol{S}]$, $\boldsymbol{Y} \sim W_r[n_2, \boldsymbol{S}]$ and $n_1, n_2 > r - 1$. Put $\boldsymbol{X} + \boldsymbol{Y} = \boldsymbol{T}'\boldsymbol{T}$, where $\boldsymbol{T}$ is the upper triangular $r \times r$ matrix of the Cholesky decomposition of $\boldsymbol{X} + \boldsymbol{Y}$. Let $\boldsymbol{P}$ be the $r \times r$ SPD random matrix defined by $\boldsymbol{X} = \boldsymbol{T}'\boldsymbol{P}\boldsymbol{T}$.*

Then, $\boldsymbol{X}+\boldsymbol{Y}$ *and* $\boldsymbol{P}$ *are independent;* $\boldsymbol{X}+\boldsymbol{Y} \sim W_r[n_1+n_2, \boldsymbol{S}]$ *and* $\boldsymbol{P} \sim$ *Beta*$_r[n_1/2, n_2/2]$.

It follows that if $\boldsymbol{X}+\boldsymbol{Y} \sim \mathrm{W}_r[n_1+n_2, \boldsymbol{S}]$ and $\boldsymbol{P} \sim \mathrm{Beta}_r[n_1/2, n_2/2]$ are independently distributed, with $\boldsymbol{X}+\boldsymbol{Y} = \boldsymbol{T}'\boldsymbol{T}$ and $\boldsymbol{X} = \boldsymbol{T}'\boldsymbol{P}\boldsymbol{T}$, then $\boldsymbol{X} \sim \mathrm{W}_r[n_1, \boldsymbol{S}]$ and $\boldsymbol{Y} \sim \mathrm{W}_r[n_2, \boldsymbol{S}]$. For more details on this, see [14, chapter 5].

Write $\boldsymbol{\Sigma}_{t-1}^{-1} = \boldsymbol{\Phi}_{t-1}$ to be the precision matrix at time $t-1$. Given the posterior

$$(\boldsymbol{\Sigma}_{t-1}|D_{t-1}) \sim \mathrm{GW}^{-1}[\boldsymbol{S}_{t-1}, \boldsymbol{N}_{t-1}, m_{t-1}],$$

for some known quantities $\boldsymbol{S}_{t-1}$ and $\boldsymbol{N}_{t-1}$, the posterior for the precision at $t-1$ is

$$(\boldsymbol{\Phi}_{t-1}|D_{t-1}) \sim \mathrm{GW}[\boldsymbol{S}_{t-1}, \boldsymbol{N}_{t-1}, m_{t-1}-r-1].$$

This can be written in the usual Wishart form as

$$(\boldsymbol{\Phi}_{t-1}|D_{t-1}) \sim \mathrm{W}_r\left[r + \frac{\mathrm{trace}\,(\boldsymbol{N}_{t-1})}{r} - 1, \boldsymbol{N}_{t-1}^{-1/2}\boldsymbol{S}_{t-1}^{-1}\boldsymbol{N}_{t-1}^{-1/2}\right]. \tag{7.14}$$

Now with $\boldsymbol{T}_{t-1}$ as the upper triangular matrix of the Cholesky decomposition of $\boldsymbol{\Phi}_{t-1}$, that is $\boldsymbol{\Phi}_{t-1} = \boldsymbol{T}_{t-1}'\boldsymbol{T}_{t-1}$, consider the resulting precision matrix

$$\boldsymbol{B}^{1/2}\boldsymbol{\Phi}_t\boldsymbol{B}^{1/2} = \boldsymbol{T}_{t-1}'\boldsymbol{P}_t\boldsymbol{T}_{t-1},$$

for an $r \times r$ symmetric positive definite matrix $\boldsymbol{P}_t$ such that

$$(\boldsymbol{\Phi}_t|D_{t-1}) \sim \mathrm{GW}[\boldsymbol{S}_{t-1}, \boldsymbol{B}\boldsymbol{N}_{t-1}, m_{B,t-1}-r-1],$$

where $\boldsymbol{B} = \mathrm{diag}\{\beta_1, \ldots, \beta_r\}$ is a discount matrix and $m_{B,t-1} = r + \mathrm{trace}\,(\boldsymbol{B}\boldsymbol{N}_{t-1})/2r$. Such a construction is always possible as shown for the Wishart case in [35].

Write again the last distribution with the Wishart notation

$$(\Phi_t|D_{t-1}) \sim \mathrm{W}_r\left[r + \frac{\mathrm{trace}\,(\boldsymbol{B}\boldsymbol{N}_{t-1})}{r} - 1, \boldsymbol{B}^{-1/2}\boldsymbol{N}_{t-1}^{-1/2}\boldsymbol{S}_{t-1}^{-1}\boldsymbol{N}_{t-1}^{-1/2}\boldsymbol{B}^{-1/2}\right],$$

so that

$$(\boldsymbol{B}^{1/2}\Phi_t\boldsymbol{B}^{1/2}|D_{t-1}) \sim \mathrm{W}_r\left[r + \frac{\mathrm{trace}\,(\boldsymbol{B}\boldsymbol{N}_{t-1})}{r} - 1, \boldsymbol{N}_{t-1}^{-1/2}\boldsymbol{S}_{t-1}^{-1}\boldsymbol{N}_{t-1}^{-1/2}\right]. \tag{7.15}$$

Using now Theorem 7.3 and equations (7.14), (7.15) it follows that

$$(\boldsymbol{P}_t|D_{t-1}) \sim \mathrm{GB}\left[\frac{1}{2}\boldsymbol{B}\boldsymbol{N}_{t-1}, \frac{1}{2}(\boldsymbol{I} - \boldsymbol{B})\boldsymbol{N}_{t-1}\right].$$

The posterior of Φ_t at t, is calculable from Theorem 4.2 by setting $\Theta = \boldsymbol{Y}_t'$ and using

$$(\boldsymbol{Y}_t'|\Sigma_t, D_{t-1}) \sim \mathrm{N}[\boldsymbol{f}_t', Q_t, \Sigma_t],$$

$$(\Sigma_t|D_{t-1}) \sim \mathrm{GW}^{-1}[\boldsymbol{S}_{t-1}, \boldsymbol{B}\boldsymbol{N}_{t-1}, m_{B,t-1}].$$

The resulting posterior is

$$(\Sigma_t|D_t) \sim \mathrm{GW}^{-1}[\boldsymbol{S}_t, \boldsymbol{N}_t, m_t],$$

where

$$\boldsymbol{N}_t^{1/2}\boldsymbol{S}_t\boldsymbol{N}_t^{1/2} = \boldsymbol{B}^{1/2}\boldsymbol{N}_{t-1}^{1/2}\boldsymbol{S}_{t-1}\boldsymbol{N}_{t-1}^{1/2}\boldsymbol{B}^{1/2} + \boldsymbol{e}_t\boldsymbol{e}_t'/Q_t$$

$$\boldsymbol{N}_t = \boldsymbol{B}\boldsymbol{N}_{t-1} + \boldsymbol{I}.$$

All the above are summarised in the following theorem.

Theorem 7.4. *One-step forecast and posterior distributions in the ECCM (7.12), (7.12′), and (7.13) are given, for each t, as follows.*

(a) Posterior at $t-1$:

For some $\boldsymbol{m}_{t-1}$, $\boldsymbol{C}_{t-1}$, $\boldsymbol{S}_{t-1}$, *and* $\boldsymbol{N}_{t-1}$,

$$(\Theta_{t-1}, \Sigma_{t-1}|D_{t-1}) \sim NGW^{-1}[\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}, \boldsymbol{S}_{t-1}, \boldsymbol{N}_{t-1}, m_{t-1}],$$

where $m_{t-1} = r + \frac{trace(\boldsymbol{N}_{t-1})}{2r}$.

(b) Prior at t:

$$(\Theta_t, \Sigma_t|D_{t-1}) \sim NGW^{-1}[\boldsymbol{a}_t, \boldsymbol{R}_t, \boldsymbol{S}_{t-1}, \boldsymbol{B}\boldsymbol{N}_{t-1}, m_{B,t-1}],$$
$$(\boldsymbol{P}_t|D_{t-1}) \sim GB\left[\frac{1}{2}\boldsymbol{B}\boldsymbol{N}_{t-1}, \frac{1}{2}(\boldsymbol{I}-\boldsymbol{B})\boldsymbol{N}_{t-1}\right],$$

with

$$\boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1}, \qquad \boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}'_t + \boldsymbol{W}_t,$$
$$\boldsymbol{P}_t^{-1} = \boldsymbol{T}_{t-1}\boldsymbol{B}^{-1/2}\Sigma_t\boldsymbol{B}^{-1/2}\boldsymbol{T}'_{t-1},$$

where $\boldsymbol{T}_{t-1}$ *is the upper triangular matrix of the Cholesky decomposition of* Σ_{t-1}^{-1}, $m_{B,t-1} = r + trace(\boldsymbol{B}\boldsymbol{N}_{t-1})/2r$, *and* $\boldsymbol{B} = diag\{\beta_1, \ldots, \beta_r\}$ $(0 < \beta_i \leq 1; 1 \leq i \leq r)$.

(c) One-step forecast:

$$(\boldsymbol{Y}'_t|\Sigma_t, D_{t-1}) \sim N[\boldsymbol{f}'_t, Q_t, \Sigma_t],$$

with marginal

$$(\boldsymbol{Y}'_t|D_{t-1}) \sim GT[\boldsymbol{f}'_t, Q_t, \boldsymbol{S}_{t-1}, \boldsymbol{B}\boldsymbol{N}_{t-1}, p_{t-1}],$$

where

$$\boldsymbol{f}'_t = \boldsymbol{F}'_t\boldsymbol{a}_t, \qquad Q_t = \boldsymbol{F}'_t\boldsymbol{R}_t\boldsymbol{F}_t + V_t, \qquad \text{and} \qquad p_{t-1} = \frac{trace(\boldsymbol{B}\boldsymbol{N}_{t-1})}{r}.$$

(d) Posterior at t:

$$(\Theta_t, \Sigma_t | D_t) \sim NGW^{-1}[\boldsymbol{m}_t, \boldsymbol{C}_t, \boldsymbol{S}_t, \boldsymbol{N}_t, m_t],$$

where

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t \boldsymbol{e}_t', \qquad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{A}_t' Q_t,$$

$$\boldsymbol{N}_t^{1/2} \boldsymbol{S}_t \boldsymbol{N}_t^{1/2} = \boldsymbol{B}^{1/2} \boldsymbol{N}_{t-1}^{1/2} \boldsymbol{S}_{t-1} \boldsymbol{N}_{t-1}^{1/2} \boldsymbol{B}^{1/2} + \boldsymbol{e}_t \boldsymbol{e}_t' / Q_t,$$

$$\boldsymbol{N}_t = \boldsymbol{B} \boldsymbol{N}_{t-1} + \boldsymbol{I}, \qquad m_t = r + \frac{trace(\boldsymbol{N}_t)}{2r},$$

and

$$\boldsymbol{A}_t = \boldsymbol{R}_t \boldsymbol{F}_t / Q_t \qquad \text{and} \qquad \boldsymbol{e}_t = \boldsymbol{Y}_t - \boldsymbol{f}_t.$$

Missing observations are treated easily by noting

$$\boldsymbol{N}_t^{1/2} \boldsymbol{S}_t \boldsymbol{N}_t^{1/2} = \boldsymbol{B}^{1/2} \boldsymbol{N}_{t-1}^{1/2} \boldsymbol{S}_{t-1} \boldsymbol{N}_{t-1}^{1/2} \boldsymbol{B}^{1/2} + \boldsymbol{U}_t \boldsymbol{e}_t \boldsymbol{e}_t' \boldsymbol{U}_t / Q_t$$

$$\boldsymbol{N}_t = \boldsymbol{B} \boldsymbol{N}_{t-1} + \boldsymbol{U}_t,$$

where $\boldsymbol{U}_t$ is as defined in Section 6.3.1 and the moments $\boldsymbol{m}_t$, $\boldsymbol{C}_t$ are updated by equations (6.4), (6.5).

In the absence of missing observations writing $\boldsymbol{N}_t = \text{diag}\{n_{1t}, \dots, n_{rt}\}$, $\boldsymbol{B} = \text{diag}\{\beta_1, \dots, \beta_r\}$ with $0 < \beta_i < 1$, the updatings of the individual degrees of freedom are $n_{it} = \beta_i n_{i,t-1} + 1$, $(i = 1, \dots, r)$, with $n_{it} \to (1 - \beta_i)^{-1}$, hence

$$\boldsymbol{N}_t \to (\boldsymbol{I} - \boldsymbol{B})^{-1},$$

as $t \to \infty$. This shows that $\boldsymbol{S}_t$ will not degenerate as t increases.

Notice that since $\boldsymbol{B} = \text{diag}\{\beta_{1t}, \dots, \beta_{rt}\}$ and $0 < \beta_{it} < 1$, $(i = 1, \dots, r)$

$$\boldsymbol{B}^t \to \text{O}.$$

Also since $\boldsymbol{N}_t$ is diagonal

$$\boldsymbol{N}_t^{-1/2} \to (\boldsymbol{I} - \boldsymbol{B})^{1/2}.$$

Write the estimate $\boldsymbol{S}_t$ of $\boldsymbol{\Sigma}_t$ as

$$\boldsymbol{S}_t = \boldsymbol{N}_t^{-1/2} \left[\boldsymbol{B}^{t/2} \boldsymbol{N}_0^{1/2} \boldsymbol{S}_0 \boldsymbol{N}_0^{1/2} \boldsymbol{B}^{t/2} + \sum_{i=0}^{t-1} \boldsymbol{B}^{i/2} \boldsymbol{e}_{t-i} \boldsymbol{e}_{t-i}' \boldsymbol{B}^{i/2} / Q_{t-i} \right] \boldsymbol{N}_t^{-1/2},$$

which implies

$$\boldsymbol{S}_t \approx (\boldsymbol{I} - \boldsymbol{B})^{1/2} \left[\sum_{i=0}^{t-1} \boldsymbol{B}^{i/2} \boldsymbol{e}_{t-i} \boldsymbol{e}_{t-i}' \boldsymbol{B}^{i/2} / Q_{t-i} \right] (\boldsymbol{I} - \boldsymbol{B})^{1/2}.$$

The procedure extends to allow a time varying discount matrix and user variance intervention.

The above analysis is trivially extended to the case when $\boldsymbol{Y}_t$ is an $r \times s$ matrix, see page 80 and Section 6.3.1.

7.3.2 The General Multivariate DLM

Consider the model of Section 5.3 with a time varying variance $\boldsymbol{\Sigma}_t$ given by

$$\boldsymbol{Y}_t = \boldsymbol{F}_t' \boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim [0, \boldsymbol{\Sigma}_t], \tag{7.16}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim [0, \boldsymbol{W}_t], \tag{7.16$'$}$$

where the quantities $\boldsymbol{F}_t$, $\boldsymbol{G}_t$, $\boldsymbol{W}_t$ are known, and the initial distributions

$$(\boldsymbol{\theta}_0 | D_0) \sim [\boldsymbol{m}_0, \boldsymbol{C}_0],$$

$$(\text{vech}(\boldsymbol{\Sigma}_0) | D_0) \sim [\text{vech}(\boldsymbol{S}_0), \boldsymbol{V}_0],$$

are partially specified for some known quantities $\boldsymbol{m}_0$, $\boldsymbol{C}_0$, $\boldsymbol{S}_0$, and $\boldsymbol{V}_0$.

A natural way of viewing the stochastic evolution of Σ_t from time $t-1$ to t is as a random walk

$$\text{vech}(\Sigma_t) = \text{vech}(\Sigma_{t-1}) + \text{vech}(\Psi_t), \qquad \text{vech}(\Psi_t) \sim [0, \boldsymbol{Z}_t], \tag{7.17}$$

for a known $r \times r$ variance matrix $\boldsymbol{Z}_t$.

Thus given Σ_{t-1}, $\text{E}[\Sigma_t|\Sigma_{t-1}, D_{t-1}] = \Sigma_{t-1}$, and $\text{V}[\text{vech}(\Sigma_t)|\Sigma_{t-1}, D_{t-1}] = \boldsymbol{Z}_t$.

$\boldsymbol{Z}_t$ is specified via a discount matrix, $\boldsymbol{B}$, as

$$\boldsymbol{Z}_t = \boldsymbol{B}^{-1/2}\boldsymbol{V}_{t-1}\boldsymbol{B}^{-1/2} - \boldsymbol{V}_{t-1},$$

where $\boldsymbol{B} = \text{diag}\{\beta_1, \dots, \beta_r\}$, $0 < \beta_i \leq 1$, $(i = 1, \dots r)$ and $\text{V}[\text{vech}(\Sigma_{t-1})|D_{t-1}] = \boldsymbol{V}_{t-1}$.

Following [41, 42] and Section 5.3 of this thesis, the weak probability assumptions are used

$$\text{vech}(\Sigma_t - \mathcal{A}_t e_t e_t' \mathcal{A}_t') \perp_1 \boldsymbol{Y}_t | D_{t-1}, \tag{7.18}$$

$$\text{C}[\text{vech}(\Sigma_t), \text{vech}(\mathcal{A}_t e_t e_t' \mathcal{A}_t')|D_{t-1}] = \text{V}[\text{vech}(\mathcal{A}_t e_t e_t' \mathcal{A}_t')|D_{t-1}], \tag{7.19}$$

where $\mathcal{A}_t = \boldsymbol{N}_t^{-1/2}\boldsymbol{B}^{1/2}\boldsymbol{S}_{t-1}^{1/2}\boldsymbol{Q}_t^{-1/2}$, $\boldsymbol{Q}_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + \boldsymbol{S}_{t-1}$.

As proposed in [42] (see also Section 5.3) the one-step errors are modelled as

$$\text{vech}(e_t e_t') = \boldsymbol{E}_t e_t + \epsilon_t, \qquad \epsilon_t \sim [\text{vech}(\boldsymbol{Q}_t), \boldsymbol{V}_{\epsilon,t}], \tag{7.20}$$

for a known $(r(r+1)/2) \times r$ design matrix $\boldsymbol{E}_t$. The variance $\boldsymbol{V}_{\epsilon,t}$ is evaluated with a discount factor δ_ϵ, as $\boldsymbol{V}_{\epsilon,t} = (1-\delta_\epsilon)\boldsymbol{E}_t\boldsymbol{Q}_t\boldsymbol{E}_t'/\delta_\epsilon$.

Then, assuming that the posterior at time $t-1$ of Σ_{t-1} is

$$(\text{vech}(\Sigma_{t-1})|D_{t-1}) \sim [\text{vech}(\boldsymbol{S}_{t-1}), \boldsymbol{V}_{t-1}],$$

for some known S_{t-1} and V_{t-1}, the prior at t is

$$(\mathrm{vech}(\Sigma_t)|D_{t-1}) \sim [\mathrm{vech}(S_{t-1}), B^{-1/2}V_{t-1}B^{-1/2}].$$

The posterior at t is calculable using (7.18), (7.19), and (7.20) as

$$(\mathrm{vech}(\Sigma_t)|D_t) \sim [\mathrm{vech}(S_t), V_t],$$

where

$$N_t^{1/2}S_tN_t^{1/2} = N_t^{1/2}S_{t-1}N_t^{1/2} + B^{1/2}S_{t-1}^{1/2}Q_t^{-1/2}(e_te_t' - Q_t)Q_t^{-1/2}S_{t-1}^{1/2}B^{1/2},$$
$$N_t = BN_{t-1} + I$$

and

$$V_t = V_{t-1} - (H_r(\mathcal{A}_t \otimes \mathcal{A}_t)G_r)E_tQ_tE_t'(H_r(\mathcal{A}_t \otimes \mathcal{A}_t)G_r)'/\delta_\epsilon,$$

where G_r is the duplication matrix, and H_r any left inverse of it.

Noting that $N_t^{-1/2}B^{1/2} = (N_{t-1} + B^{-1})^{-1/2}$ an alternative formula for S_t is

$$\begin{aligned} S_t &= S_{t-1} + (N_{t-1} + B^{-1})^{-1/2}S_{t-1}^{1/2}Q_t^{-1/2}(e_te_t' - Q_t)Q_t^{-1/2}S_{t-1}^{1/2} \\ &\quad \times(N_{t-1} + B^{-1})^{-1/2}. \end{aligned}$$

If V_0 can be freely chosen by the user (no requirement or initial specification for V_0 applies) then

$$\lim_{t\to\infty} V_t = \mathrm{O},$$

as discussed in Section 5.3.

Note that this time the posterior distribution of Σ_t concentrates about S_t as $t \to \infty$, if assumptions (7.18), (7.19) hold and V_0 is freely chosen. If (7.19) is not true, the estimate S_t remains the same, but then $\lim_{t\to\infty} V_t > \mathrm{O}$.

By setting $\boldsymbol{B} = \boldsymbol{I}$, we have $\boldsymbol{Z}_t = \mathbf{O}$ and we gain the case of a constant $\boldsymbol{\Sigma}$.

Theorem 5.3 (Section 5.4) and its upgraded form incorporating a matrix of degrees of freedom (Section 6.3.3) are automatically true, just by replacing Σ by Σ_t. This is formally presented below.

The following weak probability assumption is employed.

$$\boldsymbol{\theta}_t - \boldsymbol{A}_t \boldsymbol{Y}_t \perp_2 \boldsymbol{Y}_t | \Sigma_t, D_{t-1}. \tag{7.21}$$

Theorem 7.5. *In the multivariate DLM (7.16), (7.16′) and (7.17) using assumptions (7.21), (7.18), (7.19), and equation (7.20), one-step forecast and posterior distributions are partially given, for each t, as follows:*

(a) Posterior at $t-1$:

For some known quantities $\boldsymbol{m}_{t-1}$, $\boldsymbol{C}_{t-1}$, $\boldsymbol{S}_{t-1}$, $\boldsymbol{V}_{t-1}$

$$(\boldsymbol{\theta}_{t-1} | D_{t-1}) \sim [\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}],$$
$$(vech(\Sigma_{t-1}) | D_{t-1}) \sim [vech(\boldsymbol{S}_{t-1}), \boldsymbol{V}_{t-1}].$$

(b) Prior at t:

$$(\boldsymbol{\theta}_t | D_{t-1}) \sim [\boldsymbol{a}_t, \boldsymbol{R}_t],$$
$$(vech(\Sigma_t) | D_{t-1}) \sim [vech(\boldsymbol{S}_{t-1}), \boldsymbol{L}_t],$$

where

$$\boldsymbol{a}_t = \boldsymbol{G}_t \boldsymbol{m}_{t-1}, \qquad \boldsymbol{R}_t = \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t' + \boldsymbol{W}_t, \qquad \boldsymbol{L}_t = \boldsymbol{B}^{-1/2} \boldsymbol{V}_{t-1} \boldsymbol{B}^{-1/2},$$

and

$$\boldsymbol{B} = diag\{\beta_1, \ldots, \beta_r\}, \qquad 0 < \beta_i \leq 1, \qquad (i = 1, \ldots, r).$$

(c) One-step forecast:

$$(\boldsymbol{Y}_t|D_{t-1}) \sim [\boldsymbol{f}_t, \boldsymbol{Q}_t],$$

where

$$\boldsymbol{f}_t = \boldsymbol{F}_t'\boldsymbol{a}_t \qquad \textit{and} \qquad \boldsymbol{Q}_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + \boldsymbol{S}_{t-1}.$$

(d) Posterior at t:

$$(\boldsymbol{\theta}_t|D_t) \sim [\boldsymbol{m}_t, \boldsymbol{C}_t],$$
$$(vech(\boldsymbol{\Sigma}_t)|D_t) \sim [vech(\boldsymbol{S}_t), \boldsymbol{V}_t],$$

where

$$\begin{aligned}
\boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{A}_t\boldsymbol{e}_t, \qquad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{Q}_t\boldsymbol{A}_t', \\
\boldsymbol{S}_t &= \boldsymbol{S}_{t-1} + (\boldsymbol{N}_{t-1} + \boldsymbol{B}^{-1})^{-1/2}\boldsymbol{S}_{t-1}^{1/2}\boldsymbol{Q}_t^{-1/2}(\boldsymbol{e}_t\boldsymbol{e}_t' - \boldsymbol{Q}_t)\boldsymbol{Q}_t^{-1/2}\boldsymbol{S}_{t-1}^{1/2} \\
&\quad \times(\boldsymbol{N}_{t-1} + \boldsymbol{B}^{-1})^{-1/2}, \\
\boldsymbol{V}_t &= \boldsymbol{V}_{t-1} - (\boldsymbol{H}_r(\boldsymbol{\mathcal{A}}_t \otimes \boldsymbol{\mathcal{A}}_t)\boldsymbol{G}_r)\boldsymbol{E}_t\boldsymbol{Q}_t\boldsymbol{E}_t'(\boldsymbol{H}_r(\boldsymbol{\mathcal{A}}_t \otimes \boldsymbol{\mathcal{A}}_t)\boldsymbol{G}_r)'/\delta_\epsilon,
\end{aligned}$$

with

$$\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{Q}_t^{-1}, \qquad \boldsymbol{e}_t = \boldsymbol{Y}_t - \boldsymbol{f}_t, \qquad \boldsymbol{N}_t = \boldsymbol{B}\boldsymbol{N}_{t-1} + \boldsymbol{I},$$

and $\boldsymbol{G}_r$, $\boldsymbol{H}_r$, $\boldsymbol{\mathcal{A}}_t$, $\boldsymbol{B}$, δ_ϵ are as defined in Section 5.3.

It follows that, within the above framework, k-step forecasting and filtering moments as developed in Sections 5.4, are easily updated in the case of time-varying observational variances.

Missing observations are treated as in Section 6.3.3. Defining $\boldsymbol{U}_t$ as in

Section 6.3.1 and using the weak probability assumptions of Theorem 7.5

$$m_t = a_t + A_t U_t e_t,$$

$$C_t = R_t - A_t U_t Q_t U_t A_t',$$

$$N_t = B N_{t-1} + U_t,$$

$$S_t = S_{t-1} + N_t^{-1/2} B^{1/2} S_{t-1}^{1/2} (Q_t^{-1/2} U_t e_t e_t' U_t Q_t^{-1/2} - U_t) S_{t-1}^{1/2} B^{1/2} N_t^{-1/2}.$$

Certain intervention techniques like outlier analysis as well as variance intervention (through a time-varying variance discount matrix B_t) are some of the important applications of the above method.

7.4 Other Practical Variance Schemes

In this section a practical approach is presented that may find an appeal to practitioners. It is considered as an alternative to Sections 7.2, 7.3.

Σ_t is likely to change over reasonable periods of time and not over every t. During these periods it is constant or approximately constant. So the estimate S_t of a constant Σ may be applied for all t in every such period.

Define in general h periods, in which of one, Σ_t is not changing. That is $\Sigma_t = \Sigma^{(i)}$ for all $t_{i-1} + 1 \leq t \leq t_i$, $(i = 1, \ldots, h)$ and $t_0 = 0$. The first period is the interval $[1, t_1]$ in which Σ is constant. So Theorem 5.2 applies (under its assumptions) and if the estimate of $\Sigma^{(1)}$ in this interval is denoted by $S_t^{(1)}$

$$n_t S_t^{(1)} = n_{t-1} S_{t-1}^{(1)} + \left(S_{t-1}^{(1)}\right)^{1/2} \left(Q_t^{(1)}\right)^{-1/2} e_t e_t' \left(Q_t^{(1)}\right)^{-1/2} \left(S_{t-1}^{(1)}\right)^{1/2},$$

$$n_t = n_{t-1} + 1,$$

where $Q_t^{(1)} = F_t' R_t F_t + S_{t-1}^{(1)}$.

Then we proceed with the second period using as the initial estimate of $\Sigma^{(2)}$ not necessarily that calculated from the first period $S_{t_1}^{(1)}$. This signifies the change of the variance estimate. The first point of time in the second period is $t_1 + 1$ and the procedure can start anew.

In general denoting by $S_t^{(i)}$ the estimate of the ith period of the constant variance matrix $\Sigma^{(i)}$, then

$$n_t S_t^{(i)} = n_{t-1} S_{t-1}^{(i)} + \left(S_{t-1}^{(i)}\right)^{1/2} \left(Q_t^{(i)}\right)^{-1/2} e_t e_t' \left(Q_t^{(i)}\right)^{-1/2} \left(S_{t-1}^{(i)}\right)^{1/2},$$

$$n_t = n_{t-1} + 1,$$

where $Q_t^{(i)} = F_t' R_t F_t + S_{t-1}^{(i)}$ and $t \in [t_{i-1} + 1, t_i]$, $(i = 1, \dots, h)$.

The initial estimate $S_{t_{i-1}}^{(i)}$ can be any variance matrix. Setting $S_{t_{i-1}}^{(i)} = S_{t_{i-1}}^{(i-1)}$, it is $\Sigma^{(i-1)} = \Sigma^{(i)}$ and periods $i-1, i$ decay into one period. Of course if this happens for every i, then for every t there is no change in Σ_t.

In practice h will be relatively small and estimated by the modeller. The procedure is as follows. First a constant variance, Σ, is considered. Then, the calculated e_t's can assess the accuracy of the forecasts f_t's. Of course there is the possibility that inaccurate forecasting is the result of mis-specification of F_t or W_t. Separate monitoring procedures may be used for this. Evidence of the change of variance Σ is needed. A simple yet effective toolkit in this direction is just the graph of the available Y_t's. Another reason to address a change in the variance may be any intervention that arises from the modellers beliefs (e.g. if Y_t is a company's monthly sales and there is the information that the government is planning to increase taxation for a certain period of time, then an increase of the variance is very possible). Then, the above routine procedure can be successfully applied and strengthen interventionist procedures, see [51, chapter 11].

7.5 Illustration

In this section an illustration is given using the variance estimate of Σ_t, S_t, of Section 7.3.2. 150 observation vectors of a bivariate series, Y_t ($t = 1, \ldots, 150$), are generated. Throughout t, the observational variance Σ_t changes periodically (every 30 points of time). Model (7.16), (7.16′) of Section 7.3.2 is used for generating Y_t with the following Σ_t, ($t = 1, \ldots, 150$).

$$\Sigma_1 = \cdots = \Sigma_{30} = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}, \quad \Sigma_{31} = \cdots = \Sigma_{60} = \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix},$$

$$\Sigma_{61} = \cdots = \Sigma_{90} = \begin{pmatrix} 8 & 1.2 \\ 1.2 & 3 \end{pmatrix}, \quad \Sigma_{91} = \cdots = \Sigma_{120} = \begin{pmatrix} 10 & 1.5 \\ 1.5 & 3 \end{pmatrix},$$

$$\Sigma_{121} = \cdots = \Sigma_{150} = \begin{pmatrix} 9 & 1.6 \\ 1.6 & 3 \end{pmatrix}.$$

The other settings are

$$F_t = F = \begin{pmatrix} 1 & 0.3 \\ 0.5 & 0 \end{pmatrix}, \quad G_t = J_2(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

$$W_t = W = \begin{pmatrix} 1 & 0.5 \\ 0 & 1 \end{pmatrix},$$

and $m_0 = (1, 1)'$, $C_0 = O$.

Table C.5 (page 265) and Figure 7.1 show the simulated series. The continuous line in the figure is the first sub-series (Series 1 on Table C.5) and the dotted line is the second sub-series (Series 2 on the table). There is a clear trend in both of the sub-series, which may indicate an increase in the values of the observational variance matrix.

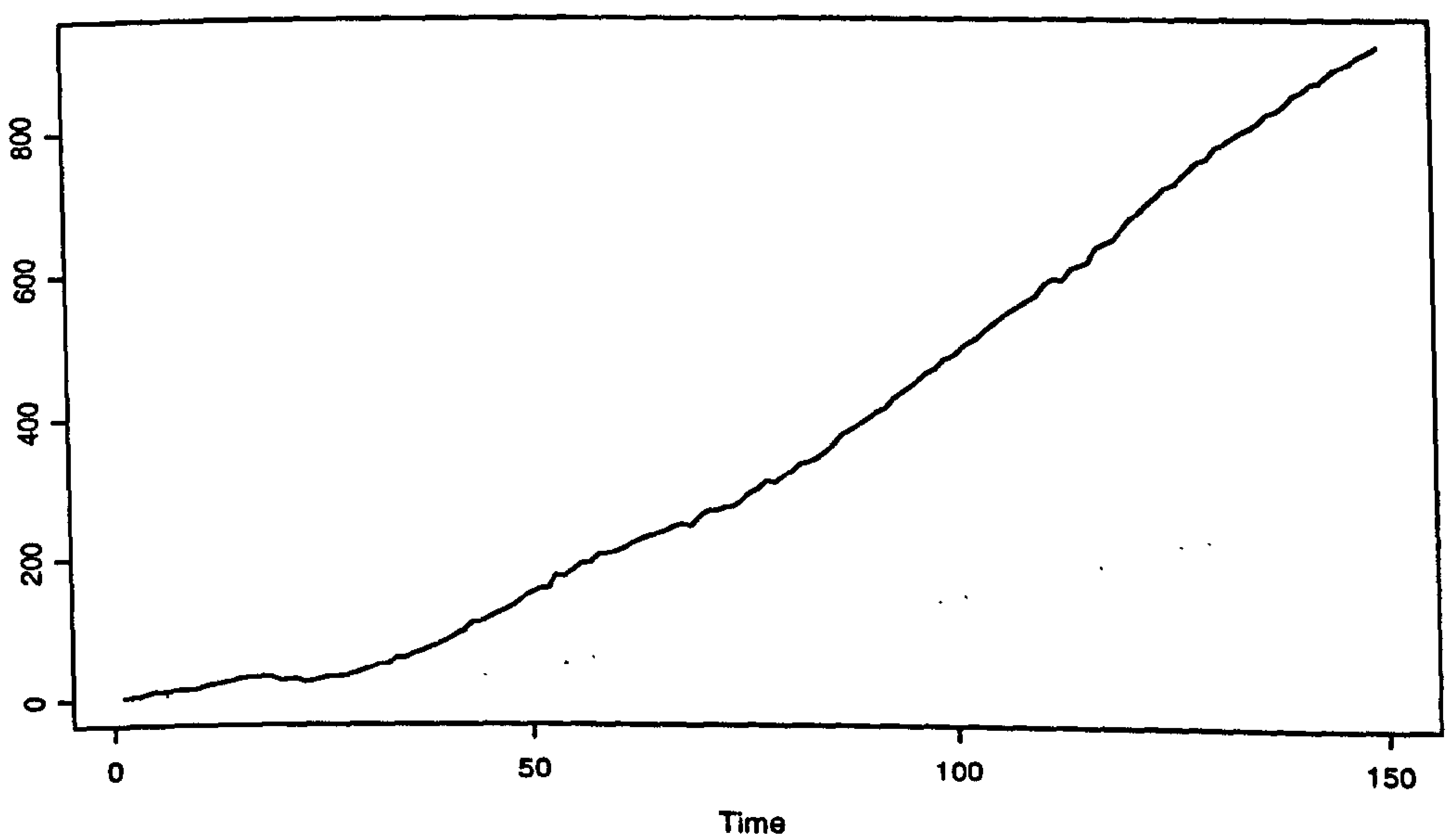

Figure 7.1: Bivariate simulated series with time-varying observational variance

The calculation of $\boldsymbol{S}_t$ is taken from Theorem 7.5 (see Section 7.3.2). The following initial quantities are used

$$\boldsymbol{N}_0 = \text{diag}\{1000^{-1}, 1000^{-1}\}, \qquad \boldsymbol{S}_0 = \text{diag}\{10, 10\}.$$

The discount matrix,

$$\boldsymbol{B} = \text{diag}\{\beta, 1\},$$

where β is a discount factor.

Write $\Sigma_t = \{\sigma_{ij,t}\}$ $(i, j = 1, 2)$. Note that the second diagonal element of $\boldsymbol{B}$ is set to 1, since the corresponding variance $\sigma_{22,t} = 3$ remains unchanged for all t.

The degrees of freedom are updated by

$$n_{1t} = \beta n_{1,t-1} + 1,$$

$$n_{2t} = n_{2,t-1} + 1,$$

where $\boldsymbol{N}_t = \text{diag}\{n_{1t}, n_{2t}\}$ is the matrix of degrees of freedom.

Write the estimate of Σ_t, $\boldsymbol{S}_t = \{s_{ij,t}\}$ $(i, j = 1, 2)$. Figures 7.2, 7.3 show the performance of the relevant estimates $s_{ij,t}$ $(i = 1, j = 1, 2)$ of $\sigma_{ij,t}$ for several values of β. The dotted lines (or grey lines) correspond to the estimate of $\sigma_{11,t}$ for $\beta = 0.7, 0.8, 0.99$ as shown from the top. The continuous line (or black line) corresponds to the estimate of σ_{11} for $\beta = 0.9$. We observe that the "black line" has a better performance of the others. The case of $\beta = 0.99$ treats the variance estimate as if Σ_t were very slowly changing ($\beta \approx 1$), while the other two ($\beta = 0.7$ and $\beta = 0.8$) treats the variance estimate as if Σ_t were very rapidly changing. For real applications if Σ_t is unknown and varies with time, it is effective that at least one discount factor be used with values around 0.97.

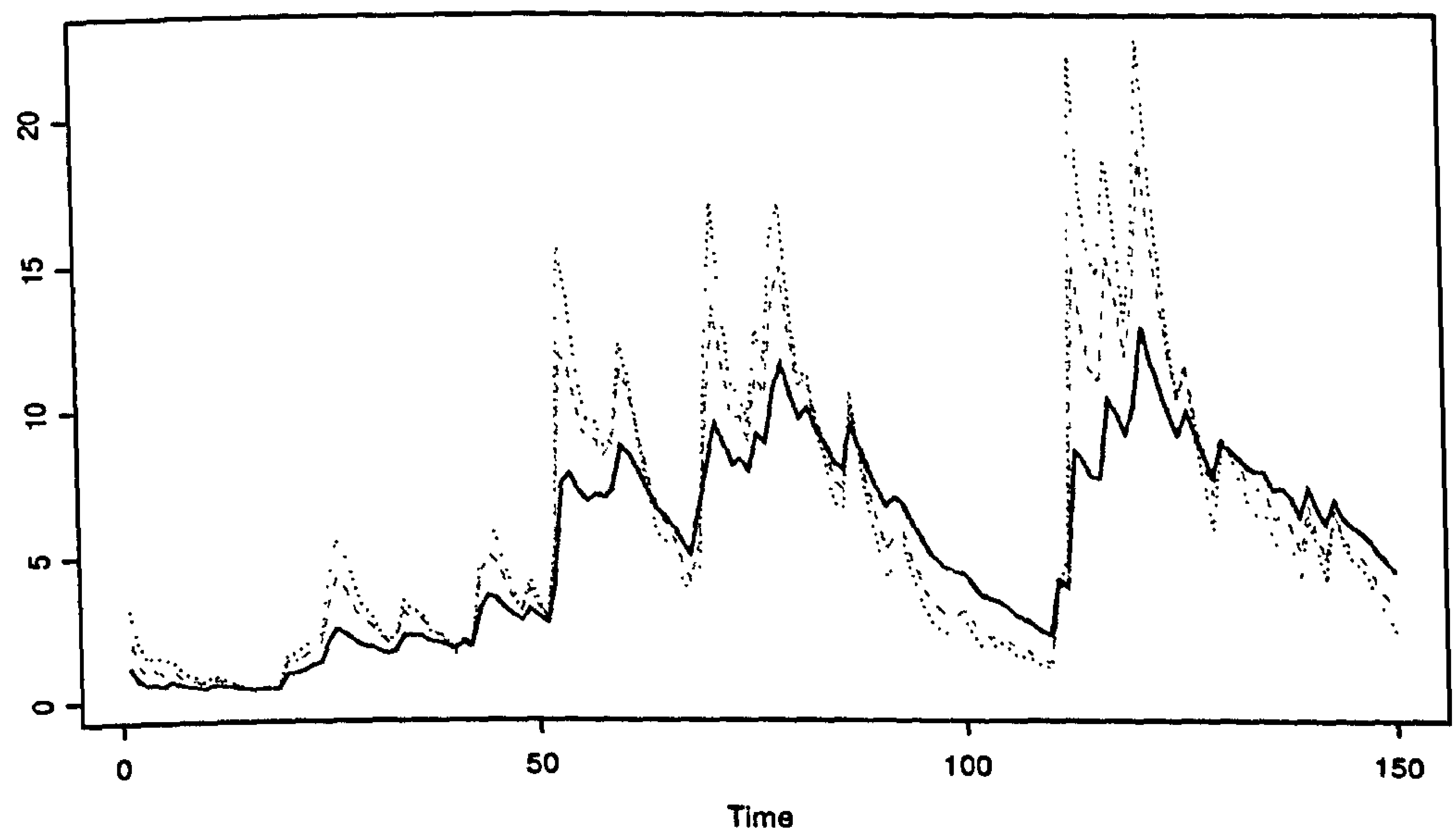

Figure 7.2: Discount factors and variance estimation for $\sigma_{11,t}$

Figure 7.4 shows the estimate of $\sigma_{22,t} = 3$. Here, the variance estimate, $s_{22,t}$, is not discounted, since $\sigma_{22,t}$ is constant throughout time.

This example illustrates the use of a diagonal matrix for variance discounting. Thus the various diagonal elements of the observational variance matrix may not be discounted at the same rate. This is particularly useful for modelling multivariate time series with different trends (e.g. see Figure 7.1). This example also illustrates the effect of different values of the discount factor β in the variance. We found that small values of $\beta = 0.7$ or $\beta = 0.8$ yield an overestimate of the corresponding variance, while a high value of $\beta = 0.99$ yields an underestimate of the variance. For this simulated series appropriate values of β may be $\beta = 0.9$ or $\beta = 0.95$.

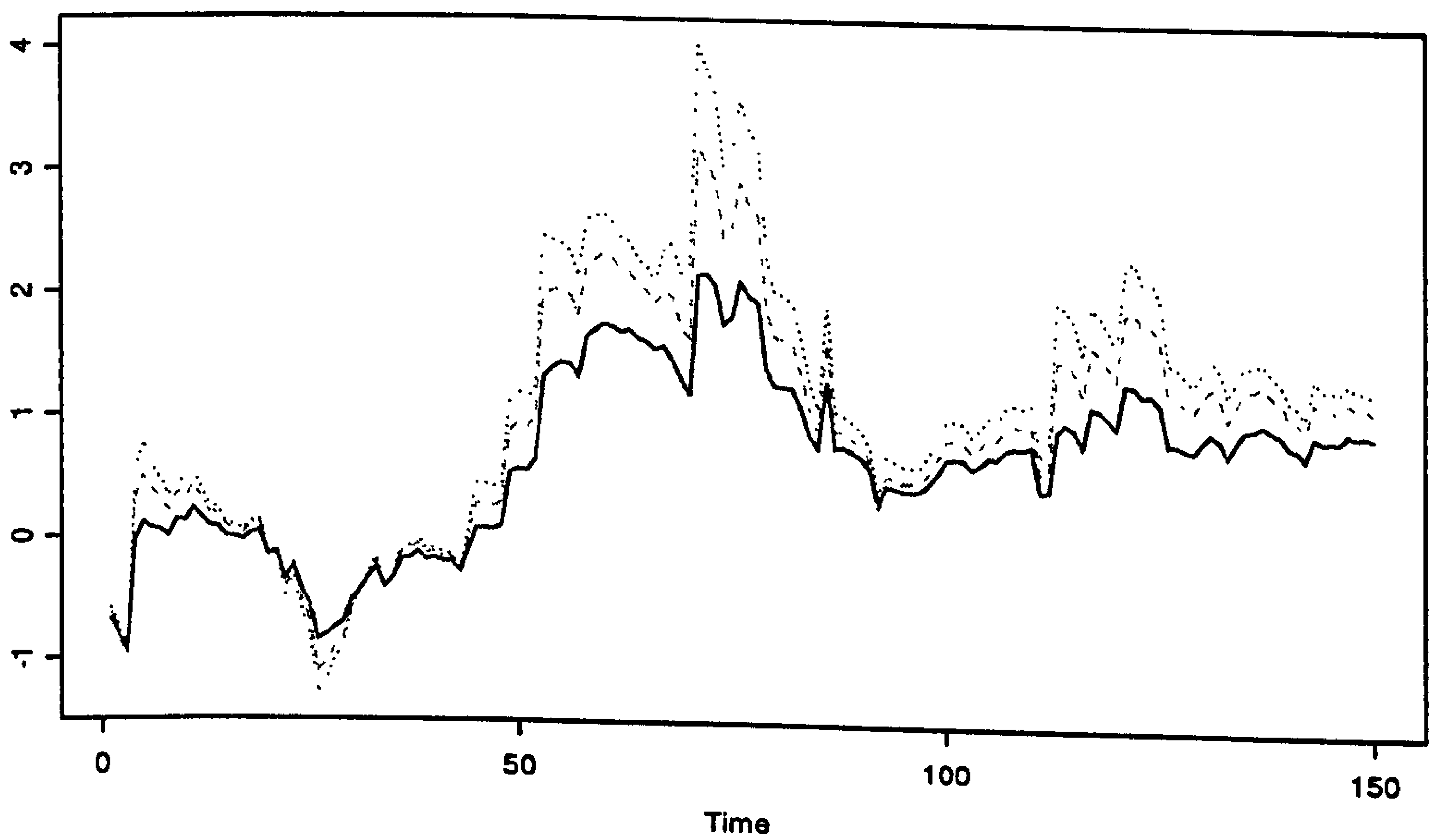

Figure 7.3: Discount factors and covariance estimation for $\sigma_{12,t}$

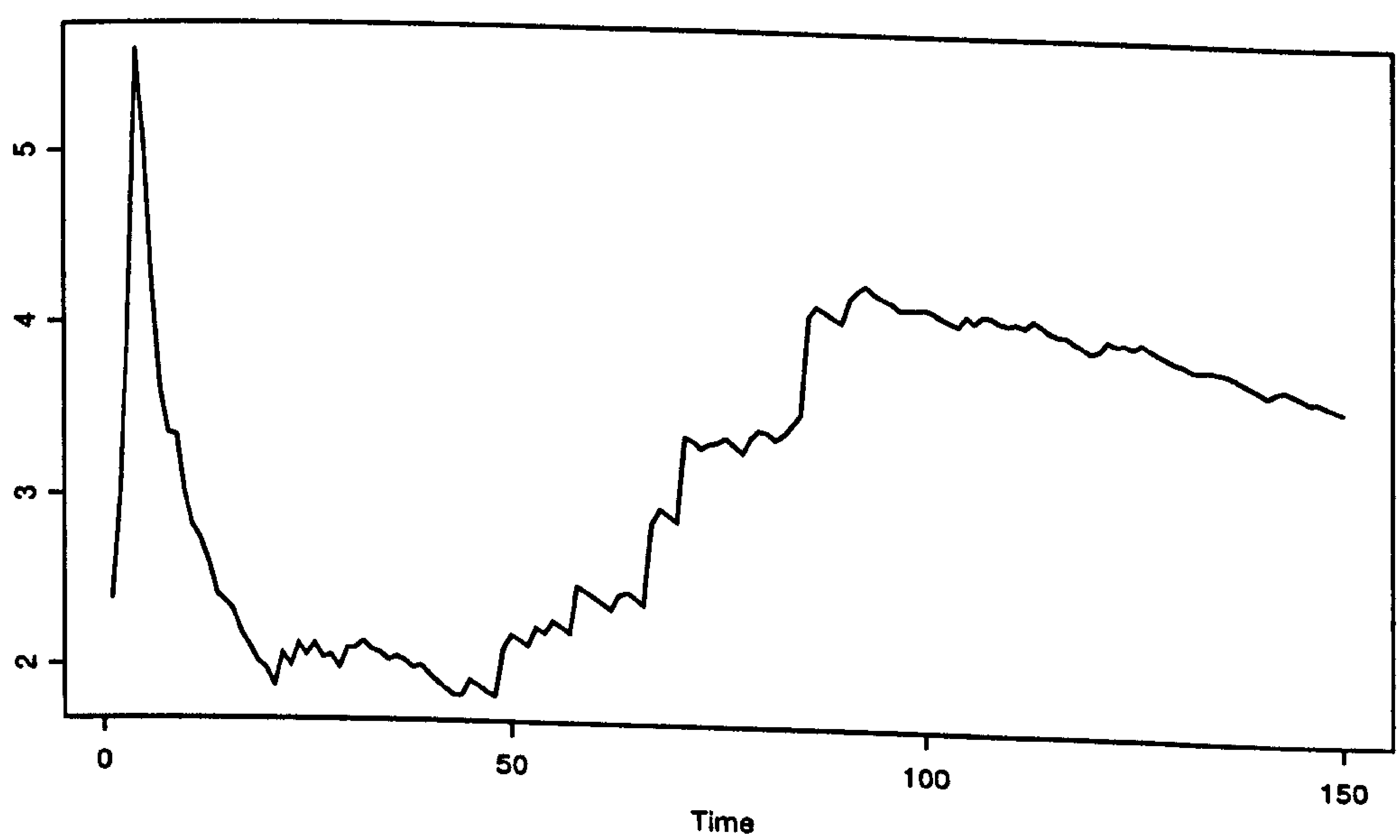

Figure 7.4: Variance estimation for $\sigma_{22,t}$

CHAPTER 8

Conclusions

8.1 Introduction

This chapter aims to summarize the methods developed in this thesis. In the following section there is a brief discussion of the achievements of the thesis.

8.2 Summary of Chapters 2-7

Chapter 2 introduced the multivariate DLM. The material is not entirely new. This is made explicit in Section 2.1. The idea was to present a unified approach to multivariate modelling rather than to compromise with modifications based on univariate models. For example observability is defined directly through the observability matrix of the multivariate model rather than through the usual definition of the linear combination of all the elements of the multivariate series, which in fact is a univariate model. Of course the result is the same, but the author believes that the preferred multivariate approach is more direct and easily understood. This led to the definitions of the canonical models in the multivariate case. This chapter serves the thesis as a background and reference chapter.

Chapter 3 dealt with a special multivariate dynamic regression model. The characteristic of this model is that no distributional assumptions about the unknown observational variance matrix was given. Thus, the Wishart limitations are avoided. The benefit of this was made explicit in Chapter 6 where the variance estimate was modified to deal with missing observations and intervention. The relationship of this new model with current matrix dynamic models is provided.

Early discussions with P.J. Harrison initiated the content of Chapter 4. These discussions were related with the introduction of the matrix of degrees of freedom. Chapter 4 has a central role in this thesis. It introduces new distributions that generalize the existing inverse Wishart, Wishart and T distributions in such a way that the scalar degrees of freedom in the latter distributions are replaced by a matrix of degrees of freedom. Several properties of these distributions are proven that are necessary for the remainder of the thesis. In the main body of the chapter the CCM and the GMDLM are extended using the above new distributions. For the ECCM (Extended CCM) the retrospective and reference analysis are developed. This generalizes a previous work, see [3]. Again the importance of these results are made evident in Chapter 6, where missing observations and related aspects are considered.

Chapter 5 encounters the problem of the general multivariate DLM with unknown observational variances. Although an approximation of this problem has previously been suggested, see [3, 4], analytic results are desirable, at least as far as the first two moments are concerned. Weak independence assumptions have been introduced, again as in Chapter 3 without specifying the distribution of the unknown variance matrix. The main results are Theorems 5.2 and 5.3. The first two moments of the filtering distributions are derived and limiting results follow in the whole of this chapter. The importance of this methodology is that not only can the modeller handle the general multivariate DLM with unknown variance matrix, but moreover the error terms distributions need not necessarily be normal. The analysis can be carried out without any loss as long as the weak assumptions hold. Applications may include univariate or multivariate models with known or unknown observational variances which are not Gaussian, or which are approximately Gaussian. Two simulations show that the method behaves significantly better than existing methods.

Missing observations in multivariate DLMs are rarely discussed in the literature. This aims to be the central subject of Chapter 6. We define partial missing observations which are virtually missing sub-vectors of the observation vector. This calls upon the distributions of Chapter 4 by individually updating each of the elements of the matrix of degrees of freedom. We provide recurrence relationships for all the models developed in Chapters 3, 4, 5, and for the multivariate model with known variances. The rest of the chapter considers the important problem of unequally spaced observations.

An application with real data shows how the new methods can be put into practice.

Chapter 7 examines the problem of time varying variances. For the general multivariate model, variance laws are considered as well as stochastic changes in the variance. For the ECCM some previous work by Quintana [33, 34] has been generalized to incorporate missing observation analysis and variance intervention. In a similar fashion as in Chapter 4 the matrix beta distribution is generalized in order to incorporate a matrix of degrees of freedom. The usual conjugacy between Wishart and beta is proven for the new distributions and a discounted matrix $\boldsymbol{B}$ replaces the usual scalar β used in the above references. This allows for a different evolution of the several elements of the variance matrix from time $t-1$ to t, whereas before all the elements of the unknown variance were discounted at the same rate. Also a methodology for the GMDLM based on the relevant theory of Chapters 5, 6 using the weak probability approach, is proposed. Again a matrix of degrees of freedom in conjunction with a matrix of discount factors allows for a very flexible variance discounting. Finally all the chapter discusses a simple and practical method for dealing with time varying variances, especially when there are not abrupt changes. In Section 7.5 a simulation demonstrates how various discount factors affect the variance estimation.

Part II

Appendices

APPENDIX A

Linear Algebra

A.1 Eigenvalues and Eigenvectors

Let $\boldsymbol{A}$ be a real valued $n \times n$ matrix. The roots of the polynomial of degree n, given by the determinant

$$p(\lambda) = |\boldsymbol{A} - \lambda \boldsymbol{I}|,$$

where $\boldsymbol{I}$ is the $n \times n$ identity matrix are called the eigenvalues of $\boldsymbol{A}$. Let $\lambda_1, \ldots, \lambda_n$ be these roots, not necessarily distinct. Then the vectors, $\boldsymbol{x}$, given by

$$\boldsymbol{A}\boldsymbol{x} = \lambda_i \boldsymbol{x},$$

for all $i = 1, \ldots, n$, are called the eigenvectors of $\boldsymbol{A}$.

Identities

(1) If $\boldsymbol{A}$ is symmetric, then λ_i are real. If in addition $\boldsymbol{A}$ is positive definite matrix, then the λ_i are positive.

(2) If $\boldsymbol{A}$ is a diagonal matrix, λ_i are its diagonal elements.

(3) If $\mathrm{rank}(\boldsymbol{A}) = n - m$, $m \leq n$, then m of the λ_i are zero.

Let now $\boldsymbol{A}$, $\boldsymbol{B}$ be any $n \times n$ matrices. Then, $\boldsymbol{A}$, $\boldsymbol{B}$ are said to be similar if there exists a non-singular $n \times n$ matrix $\boldsymbol{C}$ such that

$$\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}^{-1} = \boldsymbol{B}.$$

$\boldsymbol{C}$ is called the similarity matrix. The notation is $\boldsymbol{A} \sim \boldsymbol{B}$ and "$\sim$" is called relation of similarity or simply similarity.

Identities

(1) Both matrices $\boldsymbol{A}$, $\boldsymbol{B}$ have the same eigenvalues.

(2) The similarity is an equivalence relation.

(3) If $\boldsymbol{A}_i$, $\boldsymbol{B}_i$, are similar matrices $(i = 1, \dots, k)$ and $n = \sum_{i=1}^{k} n_i$, then the matrices defined by

$$\begin{aligned} \boldsymbol{A} &= \text{block diag}\{\boldsymbol{A}_1, \dots, \boldsymbol{A}_k\}, \\ \boldsymbol{B} &= \text{block diag}\{\boldsymbol{B}_1, \dots, \boldsymbol{B}_k\}, \end{aligned}$$

are similar, with similarity matrix

$$\boldsymbol{C} = \text{block diag}\{\boldsymbol{C}_1, \dots, \boldsymbol{C}_k\},$$

where $\boldsymbol{C}_i$ is the similarity matrix of the similar matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$, $(i = 1, \dots, k)$. $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ are said to be formed by the superposition of the matrices $\boldsymbol{A}_i$, $\boldsymbol{B}_i$, $\boldsymbol{C}_i$, $(i = 1, \dots, k)$.

If A is similar to a diagonal matrix, A is diagonalisable. Necessary and sufficient conditions for diagonalisation are given in [24, chapter 21]. The following identities are taken from this reference.

Identities

(1) A is diagonalisable, if the eigenvectors of A are linearly independent.

(2) A is diagonalisable, if its eigenvalues are distinct.

(3) A is diagonalisable, if A is symmetric or orthogonal.

Finally, we briefly discuss the Jordan forms. Let λ be any complex number and n any integer. The Jordan block, $J_n(\lambda)$ has been defined in Definition 2.6. Suppose that the $n \times n$ matrix A has $s < n$ distinct eigenvalues, λ_i, with multiplicity r_i, $(i = 1, \ldots, s)$ and $n = \sum_{i=1}^{s} r_i$. It can be shown that A is similar to the block diagonal matrix

$$J = \text{block diag}\{J_{r_1}(\lambda_1), \ldots, J_{r_s}(\lambda_s)\}.$$

Now defining the real positive powers of A as $A^k = A^{k-1}A = AA^{k-1}$, for any integer $k \geq 1$, with $A^0 = I$

$$A^k = \prod_{i=1}^{k} C^{-1}JC = C^{-1}J^kC,$$

where $A = C^{-1}JC$. Thus, matrices A^k and J^k are similar with similarity matrix C.

Consider the Jordan block with $\lambda = 0$, $J_n(0)$. Then, for any integer $k \geq n$

$$J_n^k(0) = \mathbf{O}.$$

To prove this, first note that it suffices to prove that for any $n \geq 2$ it is $J_n^n(0) = O$. Note that

$$J_n^k(0) = \begin{pmatrix} O & J_{n-k+1}(0) \\ O & O \end{pmatrix},$$

for $2 \leq k \leq n-1$. This is easily proven by induction. Then $J_n^n(0) = O$ follows immediately.

The Jordan block with $\lambda = 0$ is important. Notice that we can always write $J_n(\lambda) = \lambda I_n + J_n(0)$, for any $\lambda \in \mathbb{R}$.

A.2 Direct Product of Matrices, the Vec and Vech Operator

Let A, B be any $m \times n$ and $p \times q$ matrices respectively and write $A = \{a_{ij}\}$ $(i = 1, \ldots, m;\ j = 1, \ldots, n)$. Then, the Kronecker product or as it is sometimes referred to the direct product is denoted by $A \otimes B$ and is defined to be the $mp \times nq$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}.$$

Identities

(a) $I \otimes A = \text{block diag}\{A, \ldots, A\}$;

(b) $(A + B) \otimes C = A \otimes C + B \otimes C$,
$C \otimes (A + B) = C \otimes A + C \otimes B$;

(c) $\left(\sum_{i=1}^{r} \boldsymbol{A}_i\right) \otimes \left(\sum_{j=1}^{s} \boldsymbol{B}_j\right) = \sum_{i=1}^{r} \sum_{j=1}^{s} (\boldsymbol{A}_i \otimes \boldsymbol{B}_j)$;

(d) $(\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B}'$;

(e) $(\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C})$;

(f) $\prod_{i=1}^{k} (\boldsymbol{A}_i \otimes \boldsymbol{B}_i) = \left(\prod_{i=1}^{k} \boldsymbol{A}_i\right) \otimes \left(\prod_{i=1}^{k} \boldsymbol{B}_i\right)$;

(g) $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$;

(h) $(\boldsymbol{A} \otimes \boldsymbol{B})^{-} = \boldsymbol{A}^{-} \otimes \boldsymbol{B}^{-}$;

(i) $\mathrm{trace}(\boldsymbol{A} \otimes \boldsymbol{B}) = \mathrm{trace}(\boldsymbol{A})\mathrm{trace}(\boldsymbol{B})$;

(ia) $\mathrm{rank}(\boldsymbol{A} \otimes \boldsymbol{B}) = \mathrm{rank}(\boldsymbol{A})\mathrm{rank}(\boldsymbol{B})$,

where all matrices are appropriate in the sense that all matrix-operations are well defined and $\boldsymbol{A}^{-}$ denotes the generalized inverse of $\boldsymbol{A}$ (see Appendix A.3).

Proof. The proofs of these results can be found in [24, chapter 16]. □

Lemma A.1. *Let $\boldsymbol{A}$, $\boldsymbol{B}$ be any matrices. The matrix $\boldsymbol{A} \otimes \boldsymbol{B}$ is (a) symmetric non-negative definite if $\boldsymbol{A}$ and $\boldsymbol{B}$ are both symmetric non-negative definite or both symmetric non-positive definite, and (b) symmetric positive definite if $\boldsymbol{A}$ and $\boldsymbol{B}$ are both symmetric positive definite or both symmetric negative definite.*

Proof. See [24, chapter 16]. □

Suppose now that given an $m \times n$ matrix $\boldsymbol{A}$, $\boldsymbol{a}_i$ denotes the ith column of $\boldsymbol{A}$, $i = 1, \ldots, n$. Then $\boldsymbol{A}$ can be rearranged into an $nm \times 1$ vector,

$$\begin{pmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_n \end{pmatrix}.$$

This vector is referred to as the **vec** of $\boldsymbol{A}$ and is denoted by $\mathrm{vec}(\boldsymbol{A})$.

Identities

(a) $\mathrm{vec}\left(\sum_{i=1}^{k} c_i \boldsymbol{A}_i\right) = \sum_{i=1}^{k} c_i \mathrm{vec}(\boldsymbol{A}_i)$;

(b) $\mathrm{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{B})$,

where $\boldsymbol{A}_i$ have the same order for all $i = 1, \ldots, n$ and $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ are any matrices such that the product $\boldsymbol{ABC}$ is well defined. The proofs, again, are to be found in [24, chapter 16].

Let $\boldsymbol{A}$ be an $n \times n$ symmetric matrix. Often it is desirable that $\boldsymbol{A}$ be rearranged into a vector based upon its $n(n+1)/2$ distinctive elements (for example for differentiating purposes). If we write $\boldsymbol{A} = \{a_{ij}\}$, $(1 \leq i, j \leq n)$, then $\boldsymbol{A}$ can be rearranged into the $n(n+1)/2$ column vector $\boldsymbol{a}$, $\boldsymbol{a}' = (\boldsymbol{a}_1^{*'}, \ldots, \boldsymbol{a}_n^{*'})$, where $\boldsymbol{a}_i^* = (a_{ii}, \ldots, a_{in})$, $(1 \leq i \leq n)$. This vector is referred to as the **vech** of $\boldsymbol{A}$ and is denoted by $\mathrm{vech}(\boldsymbol{A})$.

Let $\boldsymbol{A}$ be an $n \times n$ symmetric matrix. Define the $n^2 \times n(n+1)/2$ matrix, $\boldsymbol{G}_n$, from

$$\mathrm{vec}(\boldsymbol{A}) = \boldsymbol{G}_n \mathrm{vech}(\boldsymbol{A}).$$

This matrix is called the *duplication matrix.*

It is easy to see that

$$G_1 = (1), \qquad G_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad G_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Also define the matrix $\boldsymbol{H}_n$ to be any left inverse of $\boldsymbol{G}_n$, or

$$\boldsymbol{H}_n \boldsymbol{G}_n = \boldsymbol{I}.$$

Since $\boldsymbol{G}_n$ is of full column rank, $\boldsymbol{G}_n' \boldsymbol{G}_n$ is non-singular. So one choice for $\boldsymbol{H}_n$ is

$$\boldsymbol{H}_n = (\boldsymbol{G}_n' \boldsymbol{G}_n)^{-1} \boldsymbol{G}_n'.$$

Note that it is

$$\mathrm{vech}(\boldsymbol{A}) = \boldsymbol{H}_n \mathrm{vec}(\boldsymbol{A}) \tag{A.1}$$

Let $\boldsymbol{X}$ be an $n \times n$ symmetric matrix and $\boldsymbol{A}$ an arbitrary $n \times n$ one. Then, according to [24, page 357] it is

$$\mathrm{vech}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{A}') = \boldsymbol{H}_n(\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{G}_n \mathrm{vech}(\boldsymbol{X}).$$

A.3 The Moore-Penrose Inverse

By definition any $n \times m$ matrix $\boldsymbol{G}$ that satisfies $\boldsymbol{A}\boldsymbol{G}\boldsymbol{A} = \boldsymbol{A}$, for an $m \times n$ matrix $\boldsymbol{A}$, is called the generalized inverse of $\boldsymbol{A}$. It is well known that given a matrix $\boldsymbol{G}$, there are infinitely many generalized inverses of $\boldsymbol{G}$. The next definition restricts the condition of the generalized inverse to achieve uniqueness.

Definition A.1. *An $n \times m$ matrix $\boldsymbol{G}$ is said to be the Moore-Penrose inverse or pseudoinverse of a given $m \times n$ matrix $\boldsymbol{A}$ if it satisfies*

(1) $\boldsymbol{AGA} = \boldsymbol{A}$;

(2) $\boldsymbol{GAG} = \boldsymbol{G}$;

(3) $(\boldsymbol{AG})' = \boldsymbol{AG}$;

(4) $(\boldsymbol{GA})' = \boldsymbol{GA}$.

Lemma A.2. *Let $\boldsymbol{A}$ be an $n \times n$ matrix and $\boldsymbol{P}$ an $m \times n$ matrix. If $\boldsymbol{A}$ is positive definite and $\boldsymbol{PAP}'$ is non-singular, then $\boldsymbol{PAP}'$ is also positive definite.*

Proof. See [24, chapter 14] □

Theorem A.1. *The Moore-Penrose inverse of a matrix $\boldsymbol{A}$ always exists and it is unique.*

Proof. Again, the reader is referred to [24, chapter 20]. □

The generalized inverse of a matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^-$ and the Moore-Penrose inverse of $\boldsymbol{A}$ by $\boldsymbol{A}^+$.

Theorem A.2. *For any matrix $\boldsymbol{A}$*

(i) $(\boldsymbol{A}^+)^+ = \boldsymbol{A}$;

(ii) $rank(\boldsymbol{A}^+) = rank(\boldsymbol{A})$;

(iii) If A is a non-singular $n \times n$ matrix, then $A^+ = A^{-1}$;

(iv) $(A')^+ = (A^+)'$;

(v) If A is symmetric and positive definite, then so is A^+.

Proof. See [24, chapter 20]. □

A.4 Linear Systems of Equations

Definition A.2. *Let A, B be known $m \times n$, $m \times k$ matrices, respectively. The linear system $AX = B$, where X is an unknown $n \times k$ matrix, is said to be consistent if and only if it has at least one solution. Otherwise it is called inconsistent.*

The next well known result gives necessary and sufficient conditions for consistency.

Theorem A.3. *Each of the following conditions are necessary and sufficient for a linear system $AX = B$ to be consistent.*

(i) rank(A, B) =rank(A);

(ii) $AA^- B = B$,

where A^- is a generalized inverse of A.

Proof. See [24, chapters 7, 9]. □

Corollary A.1. *If A is of full-rank, system $AX = B$ is always consistent.*

A.5 The Limit of a Sequence of Matrices

In this appendix the limit of a sequence of matrices is introduced.

Definition A.3. *The sequence of any $n \times n$ matrices $\{\boldsymbol{X}_t\}$ is said to be convergent if and only if there exists an $n \times n$ matrix $\boldsymbol{X}$ such that for all non-zero $n \times 1$ vectors $\boldsymbol{l}$, $\lim_{t\to\infty} \boldsymbol{l}'\boldsymbol{X}_t\boldsymbol{l} = \boldsymbol{l}'\boldsymbol{X}\boldsymbol{l}$.*

To the following we attempt an explanation of the above definition, which is used explicitly in this thesis. The reader may find other definitions of the limit of a sequence of matrices that would possibly lead to different limits. The purpose of this appendix is to prove the equivalence of all these definitions.

The limit of a sequence of vectors or matrices is classically introduced through the norm operator. A norm is the natural extension of the absolute value of a scalar to vectors and matrices. In the following we briefly describe this approach and prove equivalence properties.

We concentrate on the definitions related to matrices, although a matrix can always be viewed as a vector. The reason for this is that for the results throughout this thesis analytical forms of matrices are required, rather than their vectorized forms.

The definition of the norm is as follows.

Definition A.4. *The* **norm** *of a matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|$, in a linear space $\mathcal{V}$ of $m \times n$ matrices is a function $\|\cdot\| : \mathcal{V} \longrightarrow \mathbb{R}_+$ that satisfies*

(i) $\|\boldsymbol{A}\| \geq 0$ *and* $\|\boldsymbol{A}\| = 0 \Leftrightarrow \boldsymbol{A} = \mathbf{O}$;

(ii) $\|k\boldsymbol{A}\| = |k|\|\boldsymbol{A}\|$;

(iii) $\|A + B\| \leq \|A\| + \|B\|$;

(iv) $\|AB\| \leq \|A\| \|B\|$,

for A, $B \in \mathcal{V}$, $k \in \mathbb{R}$, *and the product* AB *is well defined (e.g.* $n = m$*).*

The *Eucledian norm* is

$$\|A\|_2 = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}, \tag{A.2}$$

where $A = \{a_{ij}\}$, $(i = 1, \ldots, m; j = 1, \ldots, n)$.

Definition A.5. *Suppose that* $\{X_t\}$ *be any sequence of* $n \times n$ *matrices in a linear space* $\mathcal{V}$*. Then,* $\{X_t\}$ *is said to be convergent in norm if and only if for every* $\epsilon > 0$ *there exists* $t_0 \in \mathbb{N}$ *such that for all* $t \geq t_0$, $\|X_t - X\| < \epsilon$, *for some* $n \times n$ *matrix* X. X *is called the limit matrix of* $\{X_t\}$ *and we will write* $\lim_{t \to \infty} X_t = X$.

The above definition ensures that if a limit of a sequence of matrices exists, this is unique using the same norm. The proof of this is similar to the standard proof of the uniqueness of the limit of real valued sequences. But what happens if we change norm? According to a well known theorem while the dimensions of the matrices, on which the norm is defined, are finite all the norms are equivalent. That is, they define the same topology, hence the limit is unique. This result enables us to work with a special norm, and our results will be valid for any norms (in the same topology). Next, we prove the equivalence of Definitions A.3 and A.5 using the Eucledian norm of equation (A.2). But first we need the following lemma.

Lemma A.3. *Let* $X_t = \{x_{ij,t}\}$, $X = \{x_{ij}\}$, $(i, j = 1, \ldots, n)$, *be a sequence of any* $n \times n$ *matrices. Then,* $\lim_{t \to \infty} l' X_t l$ *exists, for all* $l \in \mathbb{R}^n$, *if and only* *if* $\lim_{t\to\infty} x_{ij,t}$ *exists, for all* i,j. *In this case if* $\lim_{t\to\infty} \boldsymbol{l}'\boldsymbol{X}_t\boldsymbol{l} = \boldsymbol{l}'\boldsymbol{X}\boldsymbol{l}$, *then* $\lim_{t\to\infty} x_{ij,t} = x_{ij}$, *for all* $i,j = 1,\ldots,n$.

Proof. The proof is trivial by noticing

$$\boldsymbol{l}'\boldsymbol{X}_t\boldsymbol{l} = \sum_{i=1}^{n}\sum_{j=1}^{n} l_i l_j x_{ij,t},$$

where $\boldsymbol{l} = (l_1,\ldots,l_n)'$, and using standard limit results. □

Theorem A.4. *Let* $\{\boldsymbol{X}_t\}$ *be a sequence of* $n \times n$ *matrices in a linear space* $\mathcal{V}$. *Then the limits of Definitions A.3 and A.5 coincide.*

Proof. According to the above discussion it suffices to prove the theorem for the Eucledian norm only. Let $\boldsymbol{X}_t = \{x_{ij,t}\}$ and $\boldsymbol{X} = \{x_{ij}\}$, $(i,j = 1,\ldots,n)$.

First assume that $\boldsymbol{l}'\boldsymbol{X}_t\boldsymbol{l} \to \boldsymbol{l}'\boldsymbol{X}\boldsymbol{l}$, for a matrix $\boldsymbol{X}$ and for all $\boldsymbol{l} \in \mathbb{R}^n$. Lemma A.3 implies that for all $\epsilon_1 > 0$ there exists $t_0 > 0$ such that for every $t \geq t_0$ $|x_{ij,t} - x_{ij}| < \epsilon_1$. Now it is

$$\begin{aligned}
\|\boldsymbol{X}_t - \boldsymbol{X}\|_2 &= \left(\sum_{i=1}^{n}\sum_{j=1}^{n} |x_{ij,t} - x_{ij}|^2\right)^{1/2} \\
&\leq \sum_{i=1}^{n}\sum_{j=1}^{n} |x_{ij,t} - x_{ij}| \\
&< n^2\epsilon_1 < \epsilon,
\end{aligned}$$

if it is taken $\epsilon > n^2\epsilon_1$, and so $\boldsymbol{X}_t \to \boldsymbol{X}$ according to Definition A.5.

Now assume that for every $\epsilon > 0$ there exists $t_0 > 0$ such that for every $t \geq t_0$ $\|\boldsymbol{X}_t - \boldsymbol{X}\|_2 < \epsilon$. So $\sum_{i=1}^{n}\sum_{j=1}^{n} |x_{ij,t} - x_{ij}|^2 < \epsilon^2$, or $|x_{ij,t} - x_{ij}| < \epsilon$, which proves that $\lim_{t\to\infty} x_{ij,t} = x_{ij}$ for all i,j and from Lemma A.3 $\lim_{t\to\infty} \boldsymbol{l}'\boldsymbol{X}_t\boldsymbol{l} = \boldsymbol{l}'\boldsymbol{X}\boldsymbol{l}$. □

Let $\boldsymbol{X}$ be an $r \times r$ matrix. The spectral radius of $\boldsymbol{X}$, $\rho(\boldsymbol{X})$, is defined to be the maximum of the absolute value of all the eigenvalues of $\boldsymbol{X}$.

Lemma A.4. *Let X be an $r \times r$ matrix. If there is a matrix norm $\|\cdot\|$ such that $\|X\| < 1$, then $\lim_{n\to\infty} X^n = O$.*

Proof. See [25, chapter 5]. $\square$

Theorem A.5. *Let X be an $r \times r$ matrix. Then $\lim_{n\to\infty} X^n = O$ if and only if $\rho(X) < 1$.*

Proof. See [25, chapter 5]. $\square$

This result together with the identity

$$(I - X^n) = (I - X)(I + \cdots + X^{n-1})$$

proves that

$$\sum_{n=0}^{\infty} X^n = (I - X)^{-1}, \tag{A.3}$$

provided that all eigenvalues of X, λ_i, $(1 \le i \le r)$, satisfy $|\lambda_i| < 1$.

Definition A.5 is trivially extended to sequences of $n \times r$ matrices, since matrix norms are defined on any linear space of $n \times r$ matrices, (see Definition A.4).

More details on matrix norms can be found in [24, chapter 6] and [25, chapter 5].

A sequence of $r \times r$ matrices, $\{X_t\}$, is said to be decreasing, if and only if for all vectors $l \in \mathbb{R}^r$ and for all $t > 0$, it is $l'X_tl \ge l'X_{t+1}l$. Similarly, $\{X_t\}$ is said to be increasing if and only if for all vectors $l \in \mathbb{R}^r$ and for all $t > 0$, it is $l'X_tl \le l'X_{t+1}l$. $\{X_t\}$ is called monotonic if it is either increasing or decreasing. $\{X_t\}$ is said to be bounded if and only if exist $r \times r$ matrices M_1, M_2 such that $l'M_1l \le l'X_tl \le l'M_2l$, for all $l \in \mathbb{R}^r$ and $t \ge 0$. We shall write $M_1 \le X_t \le M_2$. If $M_1 \le X_t$, $\{X_t\}$ is said to be bounded below, while if it is $X_t \leq M_2$, $\{X_t\}$ is said to be bounded above. All variance matrices are bounded below by $\mathbf{O}$. From the above discussion it follows that if a sequence of matrices $\{X_t\}$ is monotonic and bounded, then its limit, $\lim_{t\to\infty} X_t$, exists.

A.6 Matrix Differentiation

This appendix gives the necessary ideas and results used in the text as far as matrix differentiation is concerned and it is taken from [24].

There are two kinds of derivatives that we may meet.

(i) Partial derivatives of a function of an unrestricted or symmetric matrix

Suppose that the domain of the function f to be differentiated is a set $\mathcal{S}$ in $\mathbb{R}^{m\times n}$ (the set of all real valued $m \times n$ matrices) that contains at least some interior points. Then f can be regarded as a function of an $m\times n$ matrix $X = \{x_{ij}\}$ of mn independent variables and is denoted by $f(X)$ or simply f. By definition the elements $\partial f/\partial x_{ij}$ ($i = 1, \ldots, m$; $j = 1, \ldots, n$) form a matrix, which is denoted by $\partial f(X)/\partial X$ or simply $\partial f/\partial X$ and is called the derivative of f with respect to X. Further, let us write $\partial f(X)/\partial X'$ for the $n \times m$ matrix $[\partial f(X)/X]'$ and refer to this matrix as the derivative of $f(X)$ with respect to X'. When X is a symmetric matrix the above development is no longer appropriate since there are only $m(m+1)/2$ independent variables out of the total m^2 ones. Suppose that $\mathcal{S}^*$ is the set of $m(m+1)/2$-dimensional column vectors obtained by transforming the $m\times m$ matrices in $\mathcal{S}$ from X values into $x = \mathrm{vech}(X)$ values and also suppose that $\mathcal{S}^*$ contains at least some interior points. Now, by definition the elements $\partial f/\partial x_{ij}=\partial f/\partial x_{ji}$

form a symmetric matrix that is denoted by $\partial f / \partial \boldsymbol{X}$ and is called the derivative of f with respect to $\boldsymbol{X}$.

(ii) Differentiation of matrix of functions

Suppose that there is a $p \times q$ matrix $\boldsymbol{F} = \{f_{st}\}$, $(s = 1, \ldots, p; t = 1, \ldots, q)$ of pq functions to be differentiated and that the domain of all of these functions is a set $\mathcal{S}$ in $\mathbb{R}^m$ that contains at least some interior points. $\boldsymbol{F}$ can be regarded as a function of a vector $\boldsymbol{x} = (x_1, \ldots, x_m)'$ of m independent variables. By definition all the pq elements $\partial f_{st}(\boldsymbol{x})/\partial x_j$ for all $j = 1, \ldots, m$ form a $p \times q$ matrix denoted by $\partial \boldsymbol{F}(\boldsymbol{x})/\partial x_j$ and is referred to as the jth partial derivative of $\boldsymbol{F}$ with respect to x_j, $j = 1, \ldots, m$.

Most of the results of vector real analysis can be expressed in term of matrices. The reader is referred to [24, chapter 15]. Below there are some results relevant to our study.

Given any vector valued function $\boldsymbol{F} : \mathbb{R}^m \longrightarrow \mathbb{R}^n$, the Jacobian matrix of the transformation $\boldsymbol{Y} = \boldsymbol{F}(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1, \ldots, x_m)'$ is defined to be the $m \times n$ matrix

$$\boldsymbol{J} = \frac{\partial \boldsymbol{F}(\boldsymbol{x})}{\partial \boldsymbol{x}'} = \begin{pmatrix} \partial f_1/\partial x_1 & \cdots & \partial f_1/\partial x_m \\ \vdots & \ddots & \vdots \\ \partial f_n/\partial x_1 & \cdots & \partial f_n/\partial x_m \end{pmatrix},$$

where $\boldsymbol{F} = (f_1, \ldots, f_n)'$ for some real valued functions $f_j : \mathbb{R}^m \longrightarrow \mathbb{R}$, $(j = 1, \ldots, n)$.

The determinant of $\boldsymbol{J}$ is called the Jacobian determinant or simply the Jacobian of $\boldsymbol{F}$ with respect to $\boldsymbol{x}$ and is denoted by $\mathrm{J}(\boldsymbol{x} \rightarrow \boldsymbol{F})$.

Let now $\boldsymbol{F} : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}^{k\times l}$ be a matrix valued function. Then, the Jacobian matrix of $\boldsymbol{F}$ with respect to the $m \times n$ matrix of variables $\boldsymbol{X}$, is defined to be the matrix

$$\boldsymbol{J} = \frac{\partial \text{vec}(\boldsymbol{F})}{\partial (\text{vec}\boldsymbol{X})'}.$$

The determinant of $\boldsymbol{J}$ is called the Jacobian determinant or simply the Jacobian of $\boldsymbol{F}$ with respect to $\boldsymbol{X}$. The notation is $\mathrm{J}(\boldsymbol{X} \to \boldsymbol{F})$. If $\boldsymbol{F}(\boldsymbol{X})$ is symmetric matrix, then all the above remain the same, replacing "vec" by "vech".

Now we show that if $\boldsymbol{X}$ is an $r \times r$ SPD matrix then

$$\left|\frac{\partial \text{vech}(\boldsymbol{X}^{-1})}{\partial (\text{vech}\boldsymbol{X})'}\right| = (-1)^{r(r+1)/2} |\boldsymbol{X}|^{-(r+1)}. \tag{A.4}$$

Lemma A.5. *For any $n \times n$ matrix, $\boldsymbol{A}$, using the notation of Appendix A.2*

$$|\boldsymbol{H}_n(\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{G}_n| = |\boldsymbol{A}|^{n+1}.$$

Proof. See [24, chapter 15]. □

Let us consider an $r \times r$ symmetric matrix of functions, $\boldsymbol{F}$, defined on a set $\mathcal{S}$ of a vector $\boldsymbol{x} = (x_1, \dots, x_m)'$ of m variables with $\boldsymbol{F}(\boldsymbol{x})$ be non singular $\forall \boldsymbol{x} \in \mathcal{S}$. Then

$$\frac{\partial \text{vec}(\boldsymbol{F}^{-1})}{\partial x_j} = -[\boldsymbol{F}^{-1} \otimes \boldsymbol{F}^{-1}]\frac{\partial \text{vec}(\boldsymbol{F})}{\partial x_j},$$

for every $j = 1, \dots, m$. For the proof see [24, chapter 16]. The last equation implies

$$\frac{\partial \text{vec}(\boldsymbol{F}^{-1})}{\partial \boldsymbol{x}'} = -[\boldsymbol{F}^{-1} \otimes \boldsymbol{F}^{-1}]\frac{\partial \text{vec}(\boldsymbol{F})}{\partial \boldsymbol{x}'}. \tag{A.5}$$

Now from equation (A.1) we have

$$\frac{\partial \text{vech}(\boldsymbol{F}^{-1})}{\partial \boldsymbol{x}'} = \boldsymbol{H}_r \frac{\partial \text{vec}(\boldsymbol{F}^{-1})}{\partial \boldsymbol{x}'} \tag{A.6}$$

and

$$\frac{\partial \text{vec}(\boldsymbol{F})}{\partial \boldsymbol{x}'} = \boldsymbol{G}_r \frac{\partial \text{vech}(\boldsymbol{F})}{\partial \boldsymbol{x}'}. \tag{A.7}$$

From equations (A.5), (A.6), and (A.7)

$$\frac{\partial \text{vech}(\boldsymbol{F}^{-1})}{\partial \boldsymbol{x}'} = -\boldsymbol{H}_r[\boldsymbol{F}^{-1} \otimes \boldsymbol{F}^{-1}]\boldsymbol{G}_r \frac{\partial \text{vech}\boldsymbol{F}}{\partial \boldsymbol{x}'}.$$

If we set $m = r(r+1)/2$ and use Lemma A.5

$$\left|\frac{\partial \text{vech}(\boldsymbol{F}^{-1})}{\partial \boldsymbol{x}'}\right| = (-1)^{r(r+1)/2}|\boldsymbol{F}|^{-(r+1)}\left|\frac{\partial \text{vech}\boldsymbol{F}}{\partial \boldsymbol{x}'}\right|.$$

It is just a matter of setting $\boldsymbol{x} = \text{vech}(\boldsymbol{X})$ and $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{X}$ to take equation (A.4).

Now let $\boldsymbol{A}$ be any $n \times n$ matrix, $\boldsymbol{B}$ any $r \times r$ matrix, and $\boldsymbol{X}$ an $n \times r$ matrix with variables. The Jacobian matrix of the transformation $\boldsymbol{Y} = \boldsymbol{AXB}$ is

$$\frac{\partial \text{vec}(\boldsymbol{AXB})}{\partial (\text{vec}\boldsymbol{X})'} = \boldsymbol{B}' \otimes \boldsymbol{A},$$

so that

$$\text{J}(\boldsymbol{X} \to \boldsymbol{Y}) = \left|\frac{\partial \text{vec}(\boldsymbol{AXB})}{\partial (\text{vec}\boldsymbol{X})'}\right| = |\boldsymbol{A}|^r|\boldsymbol{B}|^n.$$

If we set $\boldsymbol{B} = \boldsymbol{I}$ we get

$$\text{J}(\boldsymbol{X} \to \boldsymbol{Y}) = \left|\frac{\partial \text{vec}(\boldsymbol{AX})}{\partial (\text{vec}\boldsymbol{X})'}\right| = |\boldsymbol{A}|^r. \tag{A.8}$$

A.7 The Powers of Symmetric Matrices

The following theorem is a well known result of matrix algebra.

Theorem A.6 (Spectral Decomposition Theorem). *Let $\boldsymbol{A}$ be an $n \times n$ symmetric matrix. Then, $\boldsymbol{A}$ can be written as*

$$\boldsymbol{A} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}' = \sum_{i=1}^{n} \lambda_i \boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}'_{(i)},$$

where Λ is a diagonal matrix of the n distinct eigenvalues of A, and $\Gamma = \{\gamma_{(i)}\}$, $1 \leq i \leq n$ is a matrix whose columns correspond to the eigenvalues standardized eigenvectors.

Proof. See [28] or [24, chapter 21]. $\square$

Now we can define the rational powers of A as

$$A^{r/s} = \Gamma\Lambda^{r/s}\Gamma', \qquad \text{where} \qquad \Lambda^{r/s} = \text{diag}(\lambda_i^{r/s}),$$

for $s \in \mathbb{N}^*$ and $r \in Z$. If some of the eigenvalues of A are zero we can define only positive rational powers of A. Important special cases include $r = 1$, $s = 2$ (symmetric square root) and $r = -1$, $s = 2$. For the latter we must have $\lambda_i > 0$ for all i. If A is an SPD (symmetric positive definite) matrix it is well known that all its eigenvalues are positive real numbers (see Appendix A.1), hence there is no problem on taking any rational power. The following identities are trivially obtained

$$\begin{aligned}
A^{r_1/s_1}A^{r_2/s_2} &= A^{r_1/s_1+r_2/s_2}, &\quad& \forall s_1, s_2 \in \mathbb{N}^* &\quad& \text{and} &\quad& r_1, r_2 \in Z,\\
(A^{r/s})^m &= A^{rm/s}, && \forall s \in \mathbb{N}^* && \text{and} && r, m \in Z,\\
A^0 &= I.
\end{aligned}$$

Remarks :

(1) Theorem A.6 shows that a symmetric matrix A is uniquely identified by its distinct eigenvalues and corresponding eigenvectors.

(2) Since $A^{1/2}$ has the same eigenvectors as A and has eigenvalues which are given functions of the eigenvalues of A we see that the symmetric square root is uniquely defined.

(3) Note that in Theorem A.6 it follows that matrix $\boldsymbol{\Gamma}$ is an orthogonal matrix, or $\boldsymbol{\Gamma\Gamma'} = \boldsymbol{\Gamma'\Gamma} = \boldsymbol{I}$.

(4) From the above discussion it is clear that the power of a symmetric matrix is a symmetric matrix too.

(5) The Moore-Penrose inverse of $\boldsymbol{A}$, $\boldsymbol{A}^+$, is easily calculated by

$$\boldsymbol{A}^+ = \boldsymbol{\Gamma\Lambda^-\Gamma'},$$

where the $n \times n$ generalized inverse of $\boldsymbol{\Lambda}$, $\boldsymbol{\Lambda}^-$, is a diagonal matrix with diagonal elements $1/\lambda_i$ for all positive eigenvalues λ_i of $\boldsymbol{A}$ $(1 \le i \le n)$ and 0 for the zero eigenvalues of $\boldsymbol{A}$.

As an example consider the 2×2 matrix $\boldsymbol{A}$

$$\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}.$$

According to the above, the symmetric square root of $\boldsymbol{A}$ is

$$\boldsymbol{A}^{1/2} = \boldsymbol{\Gamma\Lambda}^{1/2}\boldsymbol{\Gamma'} = \begin{pmatrix} 1.376 & 0.325 \\ 0.325 & 1.701 \end{pmatrix}.$$

Another way of obtaining the square root of an SPD matrix is via the *Cholesky decomposition.* The next theorem provides the main result.

Theorem A.7 (Cholesky Decomposition Theorem). *For any symmetric positive definite matrix* $\boldsymbol{A}$ *there exists a unique upper triangular matrix* $\boldsymbol{T}$ *with positive diagonal elements such that*

$$\boldsymbol{A} = \boldsymbol{T'T}.$$

Then, the square root of $\boldsymbol{A}$ may be defined to be $\boldsymbol{T}$ or $\boldsymbol{T}'$.

For example, let $\boldsymbol{A}$ and $\boldsymbol{T}$ be given by

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}, \qquad \boldsymbol{T} = \begin{pmatrix} t_{11} & t_{12} \\ 0 & t_{22} \end{pmatrix},$$

where $a_{ii}, t_{ii} > 0$, $(i = 1, 2)$.

Then, by equating $\boldsymbol{T}'\boldsymbol{T} = \boldsymbol{A}$ we have

$$\boldsymbol{T} = \begin{pmatrix} a_{11}^{1/2} & a_{12}a_{11}^{-1/2} \\ 0 & (a_{22} - a_{12}^2 a_{11}^{-1})^{1/2} \end{pmatrix},$$

the square root of $\boldsymbol{A}$ based on the Cholesky decomposition.

Note that if $\boldsymbol{A}$ is diagonal matrix the symmetric square root and the square root based on the Cholesky decomposition coincide. However, it is evident that the first method is capable of providing any rational power of an SPD matrix, while the latter method provides only the square root. In this thesis the symmetric square root was preferred. The Cholesky decomposition is only used in Theorem 7.4, where this decomposition is needed to build the recurrence relationships.

APPENDIX B

Probability

B.1 Moments of Matrix Distributions

Definition B.1. *Let $\boldsymbol{X}$ be an $n \times r$ matrix. Then, $\boldsymbol{X}$ is said to have a matrix distribution with mean an $n \times r$ matrix $\boldsymbol{M}$, left variance an $n \times n$ SPD matrix $\boldsymbol{S}$, and right variance an $r \times r$ SPD matrix $\boldsymbol{\Sigma}$, if and only if*

$$vec(\boldsymbol{X}) \sim [vec(\boldsymbol{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{S}].$$

The notation is $\boldsymbol{X} \sim [\boldsymbol{M}, \boldsymbol{S}, \boldsymbol{\Sigma}]$.

Write

$$\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r), \qquad \boldsymbol{M} = (\boldsymbol{M}_1, \ldots, \boldsymbol{M}_r),$$

and $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$, $(i, j = 1, \ldots, r)$, where $\dim(\boldsymbol{X}_i) = \dim(\boldsymbol{M}_i) = n \times 1$, $(i = 1, \ldots, r)$.

Then, Definition B.1 implies

$$\boldsymbol{X}_i \sim [\boldsymbol{M}_i, \sigma_{ii}\boldsymbol{S}],$$

for $i = 1, \ldots, r$.

The next theorem provides the moments of a linear transformation of a random matrix.

Theorem B.1. *Let* $\boldsymbol{X}$ *be an* $n \times r$ *random matrix such that* $\boldsymbol{X} \sim [\boldsymbol{M}, \boldsymbol{S}, \Sigma]$*, for some known quantities* $\boldsymbol{M}$*,* $\boldsymbol{S}$*,* Σ*. Let* $\boldsymbol{A}$ *be any* $m \times n$ *matrix,* $\boldsymbol{B}$ *any* $r \times k$ *matrix, and* $\boldsymbol{C}$ *any* $m \times k$ *matrix. Then*

$$\boldsymbol{AXB} + \boldsymbol{C} \sim [\boldsymbol{AMB} + \boldsymbol{C}, \boldsymbol{ASA}', \boldsymbol{B}'\Sigma\boldsymbol{B}].$$

Proof. Using identities (a), (b) of the "vec" operator of Appendix A.2

$$\begin{aligned} \mathrm{E}[(\boldsymbol{B}' \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{X}) + \mathrm{vec}(\boldsymbol{C})] &= (\boldsymbol{B}' \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{M}) + \mathrm{vec}(\boldsymbol{C}) \\ &= \mathrm{E}[\mathrm{vec}(\boldsymbol{AXB} + \boldsymbol{C})]. \end{aligned}$$

Also from identities (d), (f) of the Kronecker product and Definition B.1

$$\begin{aligned} \mathrm{V}[(\boldsymbol{B}' \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{X}) + \mathrm{vec}(\boldsymbol{C})] &= (\boldsymbol{B}' \otimes \boldsymbol{A})(\Sigma \otimes \boldsymbol{S})(\boldsymbol{B} \otimes \boldsymbol{A}') \\ &= (\boldsymbol{B}'\Sigma\boldsymbol{B}) \otimes (\boldsymbol{ASA}') \\ &= \mathrm{V}[\mathrm{vec}(\boldsymbol{AXB} + \boldsymbol{C})]. \end{aligned}$$

So it is

$$\mathrm{vec}(\boldsymbol{AXB} + \boldsymbol{C}) \sim [\mathrm{vec}(\boldsymbol{AMB} + \boldsymbol{C}), (\boldsymbol{B}'\Sigma\boldsymbol{B}) \otimes (\boldsymbol{ASA}')]$$

and the result follows from Definition B.1. $\square$

B.2 The Matrix Normal Distribution

The non-singular matrix normal distribution is defined as follows. Assume that Θ is an $n \times r$ random matrix, $\boldsymbol{m}$ an $n \times r$ matrix, and the $n \times n$, $r \times r$ SPD matrices, $\boldsymbol{C}$, $\boldsymbol{\Sigma}$, are non-singular. Then it is said that Θ follows a matrix normal distribution if its density is expressible

$$p(\Theta) = k(\boldsymbol{C}, \boldsymbol{\Sigma}) exp\{-\frac{1}{2} trace\{(\Theta - \boldsymbol{m})' \boldsymbol{C}^{-1} (\Theta - \boldsymbol{m}) \boldsymbol{\Sigma}^{-1}\}\}, \qquad \text{(B.1)}$$

where

$$k(\boldsymbol{C}, \boldsymbol{\Sigma}) = (2\pi)^{-nr/2} |\boldsymbol{C}|^{-r/2} |\boldsymbol{\Sigma}|^{-n/2}.$$

For full details the reader may refer to [8] or [14, chapter 2]. Here, only some basic results are presented.

The full covariance structure of $\Theta = \{\theta_{ij}\}$, for $1 \leq i \leq n$, $1 \leq j \leq r$, is given by

$$\mathrm{C}[\theta_{ij}, \theta_{kl}] = c_{ik} \sigma_{jl},$$

where $\boldsymbol{C} = \{c_{ik}\}$, $1 \leq i, k \leq n$, and $\boldsymbol{\Sigma} = \{\sigma_{jl}\}$, $1 \leq j, l \leq r$ comes by rearranging the elements of Θ to form a vector (vec(Θ)) and noting that

$$\mathrm{vec}(\Theta) \sim \mathrm{N}[\mathrm{vec}(\boldsymbol{m}), \boldsymbol{\Sigma} \otimes \boldsymbol{C}],$$

where "vec", is the vec-operator and $\otimes$ the Kronecker direct product.

Now the interest lies on the marginal and conditional distributions of (B.1). Suppose that

$$\Theta = \begin{pmatrix} \Theta_1 \\ \Theta_2 \end{pmatrix}, \qquad \boldsymbol{m} = \begin{pmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \end{pmatrix}, \qquad \boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}'_{12} & \boldsymbol{C}_{22} \end{pmatrix},$$

where $\dim(\Theta_1) = \dim(\boldsymbol{m}_1) = k \times r$, $\dim(\Theta_2) = \dim(\boldsymbol{m}_2) = (n-k) \times r$, $\dim(\boldsymbol{C}_{11}) = k \times k$, $\dim(\boldsymbol{C}_{12}) = k \times (n-k)$, and $\dim(\boldsymbol{C}_{22}) = (n-k) \times (n-k)$, for any $1 \leq k \leq n-1$. Then, the marginal distribution of Θ_1 is

$$\Theta_1 \sim \mathrm{N}[\boldsymbol{m}_1, \boldsymbol{C}_{11}, \Sigma]$$

and the conditional distribution of Θ_2 given Θ_1 is

$$(\Theta_2 | \Theta_1) \sim \mathrm{N}[\boldsymbol{m}_{2|1}, \boldsymbol{C}_{2|1}, \Sigma], \tag{B.2}$$

where $\boldsymbol{m}_{2|1} = \boldsymbol{m}_2 + \boldsymbol{C}'_{12}\boldsymbol{C}_{11}^{-1}(\Theta_1 - \boldsymbol{m}_1)$ and $\boldsymbol{C}_{2|1} = \boldsymbol{C}_{22} - \boldsymbol{C}'_{12}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}$. Furthermore, the random matrices Θ_1 and $\Theta_2 - \boldsymbol{m}_{2|1}$ are independent. A similar result for subsets of columns of Θ may be obtained.

Now suppose that $(\Theta|\Sigma) \sim \mathrm{N}[\boldsymbol{m}, \boldsymbol{C}, \Sigma]$ and $\Sigma \sim \mathrm{GW}^{-1}[\boldsymbol{S}, \boldsymbol{N}, m]$, for some known quantities $\boldsymbol{m}$, $\boldsymbol{C}$, $\boldsymbol{S}$, and $\boldsymbol{N}$, where "GW^{-1}" denotes the generalized inverse Wishart distribution (see Chapter 4). The joint distribution of Θ and Σ, defined by $p(\Theta, \Sigma) = p(\Theta|\Sigma)p(\Sigma)$, is called the **normal generalized inverse Wishart** distribution, which is written using notation $(\Theta, \Sigma) \sim \mathrm{NGW}^{-1}[\boldsymbol{m}, \boldsymbol{C}, \boldsymbol{S}, \boldsymbol{N}, m]$.

B.3 The Matrix T Distribution

An $n \times r$ random matrix $\boldsymbol{T}$ is said to follow the matrix T distribution with k degrees of freedom, mode $\boldsymbol{M}$, left scale matrix $\boldsymbol{P}$, and right scale matrix $\boldsymbol{Q}$, if its density is expressible as

$$p(\boldsymbol{T}) = \frac{|\boldsymbol{Q}|^{(k+r-1)/2}|\boldsymbol{P}|^{-r/2}}{c(k,n,r)} \frac{1}{|\boldsymbol{Q} + (\boldsymbol{T}-\boldsymbol{M})'\boldsymbol{P}^{-1}(\boldsymbol{T}-\boldsymbol{M})|^{(k+n+r-1)/2}}, \tag{B.3}$$

where $\boldsymbol{M}$ is an $n \times r$ matrix, $\boldsymbol{P}$ an $n \times n$ SPD matrix, $\boldsymbol{Q}$ an $r \times r$ SPD matrix, k any positive real number, and

$$c(k,n,r) = \frac{\pi^{nr/2}\Gamma_r\left(\frac{k+r-1}{2}\right)}{\Gamma_r\left(\frac{k+n+r-1}{2}\right)}, \tag{B.4}$$

with

$$\Gamma_r(x) = \pi^{r(r-1)/4}\Gamma(x)\Gamma(x-\frac{1}{2})\cdots\Gamma\left(x-\frac{r}{2}+\frac{1}{2}\right),$$

and $\Gamma(\cdot)$ be the gamma function.

This distribution is denoted by $\boldsymbol{T} \sim \mathrm{T}_k[\boldsymbol{M}, \boldsymbol{P}, \boldsymbol{Q}]$. Note that if $r = 1$ the above distribution reduces to a multivariate T with the same degrees of freedom and scale matrix $\boldsymbol{QP}$, namely $\boldsymbol{T} \sim \mathrm{T}_k[\boldsymbol{M}, \boldsymbol{QP}]$, where $\boldsymbol{M}$ is an $n \times 1$ vector, $\boldsymbol{P}$ an $n \times n$ SPD matrix, and $Q, k > 0$ (see [14, page 134]).

Next the marginal and conditional distributions of the matrix T distribution, are derived.

Let $\boldsymbol{T} \sim \mathrm{T}_k[\boldsymbol{M}, \boldsymbol{P}, \boldsymbol{Q}]$, as defined before, and the following partition

$$\boldsymbol{T} = (\boldsymbol{T}_1, \boldsymbol{T}_2), \qquad \boldsymbol{M} = (\boldsymbol{M}_1, \boldsymbol{M}_2), \qquad \boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}'_{12} & \boldsymbol{Q}_{22} \end{pmatrix},$$

where $\dim(\boldsymbol{T}_1) = \dim(\boldsymbol{M}_1) = n \times q$, $\dim(\boldsymbol{T}_2) = \dim(\boldsymbol{M}_2) = n \times (r-q)$, $\dim(\boldsymbol{Q}_{11}) = q \times q$, $\dim(\boldsymbol{Q}_{12}) = q \times (r-q)$, and $\dim(\boldsymbol{Q}_{22}) = (r-q) \times (r-q)$, for some $1 \leq q \leq r-1$.

Then, the marginal distribution of $\boldsymbol{T}_2$ is

$$\boldsymbol{T}_2 \sim \mathrm{T}_k[\boldsymbol{M}_2, \boldsymbol{P}, \boldsymbol{Q}_{22}] \tag{B.5}$$

and the conditional distribution of $\boldsymbol{T}_2$, given $\boldsymbol{T}_1$ is

$$(\boldsymbol{T}_2 | \boldsymbol{T}_1) \sim \mathrm{T}_{k+q}[\boldsymbol{M}_{2|1}, \boldsymbol{P}_2, \boldsymbol{Q}_{2|1}], \tag{B.6}$$

where

$$\boldsymbol{M}_{2|1} = \boldsymbol{M}_2 + (\boldsymbol{T}_1 - \boldsymbol{M}_1)\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12},$$

$$\boldsymbol{P}_2 = \boldsymbol{P} + (\boldsymbol{T}_1 - \boldsymbol{M}_1)\boldsymbol{Q}_{11}^{-1}(\boldsymbol{T}_1 - \boldsymbol{M}_1)',$$

$$\boldsymbol{Q}_{2|1} = \boldsymbol{Q}_{22} - \boldsymbol{Q}'_{12}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12},$$

provided that $\boldsymbol{Q}_{11}$ is non-singular.

A similar result may be derived for a row partition of $\boldsymbol{T}$.

More details about matrix distributions can be found in [14].

APPENDIX C

Tables

C.1 Table of Chapter 5, Section 5.6

Table C.1: 2 bivariate simulated series

time	Series a 1	Series a 2	Series b 1	Series b 2
1	-0.741	3.349	2.973	6.179
2	-1.111	1.214	3.816	5.989
3	-0.082	-0.217	4.613	7.778
4	2.333	0.602	3.203	6.611
5	4.435	7.844	6.069	11.575
6	9.839	12.900	3.290	6.175

continued on next page

continued from previous page

time	Series a 1	Series a 2	Series b 1	Series b 2
7	13.411	20.406	3.662	8.523
8	14.517	26.699	4.933	5.119
9	15.517	26.867	-1.022	3.339
10	15.429	33.715	-2.249	-2.238
11	15.770	29.588	2.208	2.053
12	24.396	39.514	-0.438	1.099
13	33.677	52.580	-2.210	-3.838
14	40.503	68.882	-4.206	-7.442
15	43.738	81.599	-9.943	-14.072
16	40.962	86.071	-10.279	-18.447
17	48.100	90.400	-19.557	-27.913
18	56.178	101.860	-21.716	-38.122
19	64.010	115.636	-28.878	-48.024
20	65.955	124.646	-35.423	-59.469
21	76.421	138.711	-40.452	-69.327
22	78.106	147.606	-48.663	-86.042
23	83.434	156.431	-55.688	-97.165
24	92.773	168.567	-63.331	-112.139
25	93.309	179.194	-67.615	-126.675
26	103.661	189.527	-76.550	-142.173
27	113.752	206.955	-80.118	-152.070
28	127.299	226.713	-84.373	-163.664

continued on next page

continued from previous page

time	Series a 1	Series a 2	Series b 1	Series b 2
29	141.062	254.583	-89.860	-171.605
30	154.986	286.670	-93.007	-177.078
31	172.299	312.882	-98.969	-186.784
32	183.257	339.999	-96.580	-189.091
33	188.461	358.279	-99.080	-193.187
34	195.405	373.534	-101.131	-197.191
35	206.772	391.778	-107.910	-205.554
36	210.950	410.502	-108.767	-212.064
37	217.618	423.453	-116.108	-222.385
38	224.688	431.972	-122.189	-239.091
39	239.117	455.603	-126.743	-248.660
40	251.722	476.411	-127.741	-248.396
41	265.228	503.161	-128.070	-253.078
42	277.619	531.151	-136.768	-262.513
43	289.271	550.596	-138.441	-269.062
44	304.712	582.441	-145.626	-280.026
45	321.145	608.226	-151.228	-287.650
46	330.498	637.494	-150.752	-294.319
47	338.197	658.477	-160.119	-304.360
48	348.038	676.450	-164.477	-316.847
49	362.866	700.754	-169.229	-328.140
50	362.734	716.254	-179.388	-345.776

continued on next page

continued from previous page

time	Series a 1	Series a 2	Series b 1	Series b 2
51	371.518	728.552	-181.678	-356.239
52	377.712	743.393	-185.632	-360.126
53	387.484	755.729	-189.731	-373.288
54	393.008	772.439	-193.058	-378.904
55	401.693	786.143	-195.544	-387.728
56	409.370	803.893	-205.187	-398.618
57	420.532	821.724	-211.148	-409.019
58	433.770	845.124	-211.979	-415.781
59	445.739	869.801	-218.334	-426.859
60	459.354	894.800	-222.651	-437.219
61	476.250	920.220	-224.878	-444.820
62	488.336	952.050	-231.388	-452.862
63	497.107	974.937	-234.883	-463.603
64	513.698	1001.155	-244.685	-476.206
65	527.531	1030.913	-246.473	-482.436
66	535.409	1051.595	-254.361	-498.065
67	544.596	1076.245	-262.114	-510.456
68	556.487	1093.973	-263.797	-521.389
69	561.961	1109.595	-274.610	-536.626
70	569.884	1122.420	-284.140	-551.240
71	576.258	1139.244	-290.158	-561.637
72	588.746	1156.374	-295.935	-577.874

continued on next page

continued from previous page

time	Series a 1	Series a 2	Series b 1	Series b 2
73	604.287	1180.739	-304.340	-594.403
74	618.669	1205.429	-308.420	-606.048
75	635.902	1237.749	-314.551	-614.902
76	646.738	1265.806	-318.643	-631.080
77	668.334	1304.090	-322.228	-634.362
78	689.315	1342.186	-327.034	-644.417
79	702.173	1374.614	-333.888	-653.636
80	716.912	1402.938	-342.977	-668.472
81	727.111	1430.111	-347.367	-685.227
82	730.859	1448.438	-350.326	-690.887
83	743.806	1469.041	-354.552	-702.157
84	756.865	1487.581	-360.070	-713.370
85	771.592	1510.529	-363.077	-718.967
86	787.208	1543.899	-370.086	-728.650
87	805.966	1577.290	-373.304	-735.430
88	825.613	1611.398	-379.286	-748.758
89	838.832	1646.875	-388.585	-764.438
90	844.053	1668.664	-394.949	-777.204
91	858.276	1694.523	-399.938	-788.843
92	870.129	1714.240	-409.563	-802.884
93	879.295	1734.999	-415.003	-815.325
94	890.883	1760.969	-420.627	-831.872

continued on next page

continued from previous page				
time	Series a 1	Series a 2	Series b 1	Series b 2
95	897.629	1782.682	-426.379	-840.542
96	906.054	1794.449	-435.739	-857.025
97	916.405	1820.222	-443.016	-875.766
98	923.886	1832.233	-452.039	-883.798
99	927.054	1843.050	-458.185	-899.960
100	930.009	1854.089	-465.453	-915.254

C.2 Tables of Chapter 6

C.2.1 Table of Section 6.3

Table C.2: Calculus of missing and observed values

(observed)+(observed)=(observed)
(observed)+(missing)=(missing)
(missing)+(observed)=(missing)
(missing)+(missing)=(missing)
(observed)×(observed)=(observed)
(observed)×(missing) =(missing)
(missing)×(observed)=(missing)
(missing) ×(missing)=(missing)

C.2.2 Tables of Section 6.5

Table C.3: London metal exchange official closing prices of aluminium in US $ per tonne

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
2000/3/2	1599.5/1600.0	1615.5/1616.0	1587.0/1592.0	1553.0/1558.0	1600.0
3	1586.0/1587.0	1606.5/1607.0	1582.0/1587.0	1552.0/1557.0	1587.0
6	1586.5/1587.5	1609.0/1610.0	1582.0/1587.0	1550.0/1555.0	1587.5
7	1595.0/1595.5	1620.0/1620.5	1590.0/1595.0	1555.0/1560.0	1595.5
8	1559.0/1560.0	1587.0/1587.5	1568.0/1573.0	1540.0/1545.0	1560.0
9	1562.5/1563.0	1588.5/1589.5	1572.0/1577.0	1545.0/1550.0	1563.0
10	1576.0/1577.0	1600.0/1601.0	1575.0/1580.0	1547.0/1552.0	1577.0
13	1570.0/1571.0	1599.0/1599.5	1575.0/1580.0	1543.0/1548.0	1571.0
14	1585.0/1586.0	1610.5/1611.0	1583.0/1588.0	1550.0/1555.0	1586.0
15	1595.0/1595.5	1618.5/1619.0	1585.0/1590.0	1547.0/1552.0	1595.5
16	1592.0/1593.0	1613.0/1613.5	1575.0/1580.0	1537.0/1542.0	1593.0
17	1600.5/1601.0	1619.0/1620.0	1582.0/1587.0	1547.0/1552.0	1601.0
20	1598.5/1599.0	1618.0/1618.5	1577.0/1582.0	1540.0/1545.0	1599.0
21	1589.0/1589.5	1614.0/1614.5	1577.0/1582.0	1545.0/1550.0	1589.5
22	1576.5/1577.0	1606.5/1607.0	1583.0/1588.0	1553.0/1558.0	1577.0
23	1568.5/1569.5	1596.0/1597.0	1582.0/1587.0	1588.0/1563.0	1569.5
24	1569.0/1570.0	1596.0/1597.0	1580.0/1585.0	1555.0/1560.0	1570.0
27	1558.0/1558.5	1588.0/1588.5	1577.0/1582.0	1553.0/1558.0	1558.5

continued on next page

continued from previous page

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
28	1566.0/1566.5	1593.0/1593.5	1580.0/1585.0	1558.0/1563.0	1566.5
29	1555.0/1555.5	1584.0/1584.5	1573.0/1578.0	1550.0/1555.0	1555.5
30	1561.0/1561.5	1588.0/1589.0	1572.0/1577.0	1547.0/1552.0	1561.5
31	1522.5/1523.0	1549.0/1549.5	1545.0/1550.0	1528.0/1533.0	1523.0
2000/4/3	1509.0/1510.0	1536.0/1537.0	1540.0/1545.0	1515.0/1520.0	1510.0
4	1494.0/1495.0	1527.0/1528.0	1540.0/1545.0	1515.0/1520.0	1510.0
5	1479.5/1480.5	1507.0/1507.5	1525.0/1530.0	1520.0/1525.0	1480.5
6	1480.0/1480.5	1508.0/1509.0	1528.0/1533.0	1520.0/1525.0	1480.5
7	1459.0/1460.0	1491.0/1492.0	1523.0/1528.0	1520.0/1525.0	1460.0
10	1466.0/1467.0	1501.0/1501.5	1528.0/1533.0	1523.0/1528.0	1467.0
11	1473.0/1474.0	1504.0/1505.0	1530.0/1535.0	1528.0/1533.0	1474.0
12	1464.0/1464.5	1495.0/1496.0	1517.0/1522.0	1513.0/1518.0	1464.5
13	1464.0/1465.0	1496.0/1497.0	1520.0/1525.0	1513.0/1518.0	1465.0
14	1438.5/1439.5	1471.5/1472.5	1502.0/1507.0	1508.0/1513.0	1439.5
17	1396.0/1397.0	1427.0/1428.0	1480.0/1485.0	1493.0/1498.0	1397.0
18	1426.0/1427.0	1458.0/1459.0	1505.0/1510.0	1517.0/1522.0	1427.0
19	1433.0/1434.0	1466.0/1467.0	1515.0/1520.0	1528.0/1533.0	1434.0
20	1428.0/1429.0	1457.0/1458.0	1498.0/1503.0	1507.0/1512.0	1429.0
21	***	***	***	***	***
24	***	***	***	***	***
25	1441.0/1442.0	1473.0/1474.0	1513.0/1518.0	1520.0/1525.0	1442.0

continued on next page

continued from previous page

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
26	1464.0/1465.0	1495.0/1495.5	1528.0/1533.0	1530.0/1535.0	1465.0
27	1451.0/1452.0	1479.5/1480.5	1508.0/1513.0	1508.0/1513.0	1452.0
28	1454.0/1455.0	1484.0/1485.0	1512.0/1517.0	1512.0/1517.0	1455.0
2000/5/2	1452.5/1453.5	1479.0/1479.5	1505.0/1510.0	1503.0/1508.0	1453.5
3	1465.0/1466.0	1493.0/1494.0	1517.0/1522.0	1515.0/1520.0	1466.0
4	1441.0/1442.0	1468.0/1469.0	1502.0/1507.0	1505.0/1510.0	1442.0
5	1460.0/1460.5	1488.0/1489.0	1518.0/1523.0	1523.0/1528.0	1460.5
8	1445.0/1445.5	1474.0/1474.5	1503.0/1508.0	1507.0/1512.0	1445.5
9	1435.0/1436.0	1464.5/1465.0	1502.0/1507.0	1512.0/1517.0	1436.0
10	1458.0/1458.5	1486.0/1487.0	1518.0/1523.0	1525.0/1530.0	1458.5
11	1430.0/1430.5	1460.0/1461.0	1498.0/1503.0	1512.0/1517.0	1430.5
12	1453.0/1455.0	1481.0/1482.0	1515.0/1520.0	1525.0/1530.0	1455.0
15	1488.0/1448.5	1473.0/1474.0	1503.0/1508.0	1510.0/1515.0	1448.5
16	1490.0/1491.0	1511.5/1512.0	1533.0/1538.0	1538.0/1543.0	1491.0
17	1514.0/1514.5	1534.0/1534.5	1543.0/1548.0	1543.0/1548.0	1514.5
18	1512.0/1513.0	1533.0/1534.0	1543.0/1548.0	1545.0/1550.0	1513.0
19	1497.5/1498.0	1518.5/1519.5	1532.0/1537.0	1535.0/1540.0	1498.0
22	1495.5/1496.0	1518.5/1519.5	1530.0/1535.0	1567.0/1572.0	1496.0
23	1492.5/1493.5	1516.0/1517.0	1527.0/1532.0	1528.0/1533.0	1493.5
24	1479.5/1480.0	1504.0/1505.0	1520.0/1525.0	1522.0/1527.0	1480.0
25	1468.0/1469.0	1497.0/1498.0	1513.0/1518.0	1515.0/1520.0	1469.0

continued on next page

continued from previous page

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
26	1463.0/1464.0	1490.0/1491.0	1513.0/1518.0	1520.0/1525.0	1464.0
29	***	***	***	***	***
30	1455.5/1456.0	1485.5/1486.0	1508.0/1513.0	1515.0/1520.0	1456.0
31	1439.0/1440.0	1465.5/1466.0	1502.0/1507.0	1517.0/1522.0	1440.0
2000/6/1	1442.0/1443.0	1465.0/1466.0	1500.0/1505.0	1515.0/1520.0	1443.0
2	1463.0/1464.0	1476.5/1477.0	1503.0/1508.0	1513.0/1518.0	1464.0
5	1419.0/1419.5	1440.0/1441.0	1478.0/1483.0	1495.0/1500.0	1419.5
6	1443.0/1443.5	1464.0/1464.5	1500.0/1505.0	1518.0/1523.0	1443.5
7	1437.0/1437.5	1458.0/1458.5	1500.0/1505.0	1522.0/1527.0	1437.5
8	1443.0/1444.0	1463.5/1464.0	1508.0/1513.0	1530.0/1535.0	1444.0
9	1456.0/1456.5	1475.0/1476.0	1510.0/1515.0	1527.0/1532.0	1456.5
12	1457.0/1458.0	1477.0/1477.5	1508.0/1513.0	1520.0/1525.0	1458.0
13	1448.0/1449.0	1466.0/1466.5	1498.0/1503.0	1510.0/1515.0	1449.0
14	1478.0/1479.0	1493.0/1494.0	1513.0/1518.0	1520.0/1525.0	1479.0
15	1546.0/1547.0	1557.0/1558.0	1555.0/1560.0	1548.0/1553.0	1547.0
16	1554.0/1555.0	1568.0/1569.0	1560.0/1565.0	1547.0/1552.0	1555.0
19	1550.0/1551.0	1571.0/1572.0	1568.0/1573.0	1552.0/1557.0	1551.0
20	1534.0/1535.0	1554.5/1555.0	1552.0/1557.0	1543.0/1548.0	1535.0
21	1539.0/1540.0	1557.0/1558.0	1552.0/1557.0	1543.0/1548.0	1540.0
22	1550.5/1551.0	1568.5/1569.0	1558.0/1563.0	1548.0/1553.0	1551.0
23	1544.0/1544.5	1561.5/1562.0	1548.0/1553.0	1535.0/1540.0	1544.5

continued on next page

continued from previous page

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
26	1554.0/1555.0	1577.5/1578.0	1565.0/1570.0	1552.0/1557.0	1555.0
27	1579.0/1579.5	1599.0/1600.0	1583.0/1588.0	1568.0/1573.0	1579.5
28	1574.0/1575.0	1593.0/1594.0	1577.0/1582.0	1562.0/1567.0	1575.0
29	1556.0/1557.0	1576.0/1577.0	1562.0/1567.0	1548.0/1553.0	1557.0
30	1563.0/1564.0	1581.0/1582.0	1568.0/1573.0	1553.0/1558.0	1564.0
2000/7/3	1560.0/1560.5	1580.0/1580.5	1560.0/1565.0	1545.0/1550.0	1560.5
4	1552.0/1553.0	1574.0/1575.0	1555.0/1560.0	1542.0/1547.0	1553.0
5	1576.0/1568.0	1588.0/1589.0	1568.0/1573.0	1555.0/1560.0	1568.0
6	1560.0/1560.5	1583.0/1584.0	1563.0/1568.0	1550.0/1555.0	1560.5
7	1542.0/1543.0	1567.0/1568.0	1553.0/1558.0	1540.0/1545.0	1543.0
10	1540.5/1541.0	1566.0/1567.0	1553.0/1558.0	1540.0/1545.0	1541.0
11	1568.0/1569.0	1592.0/1593.0	1578.0/1583.0	1565.0/1570.0	1569.0
12	1562.5/1563.0	1587.0/1587.5	1570.0/1575.0	1558.0/1563.0	1563.0
13	1572.5/1573.5	1591.5/1592.5	1563.0/1568.0	1548.0/1553.0	1573.5
14	1561.0/1562.0	1582.0/1582.5	1553.0/1558.0	1538.0/1543.0	1562.0
17	1571.5/1572.5	1591.5/1592.0	1558.0/1563.0	1545.0/1550.0	1572.5
18	1598.0/1599.0	1617.5/1618.0	1573.0/1578.0	1558.0/1563.0	1599.0
19	1586.0/1587.0	1607.0/1608.0	1563.0/1568.0	1548.0/1553.0	1587.0
20	1562.0/1562.5	1585.0/1585.5	1550.0/1555.0	1535.0/1540.0	1562.5
21	1580.5/1581.0	1601.0/1601.5	1557.0/1562.0	1540.0/1545.0	1581.0
24	1568.0/1569.0	1589.0/1589.5	1552.0/1557.0	1537.0/1542.0	1569.0

continued on next page

continued from previous page

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
25	1567.0/1567.5	1588.0/1589.0	1545.0/1550.0	1532.0/1537.0	1567.5
26	1567.0/1568.0	1588.5/1589.0	1550.0/1555.0	1535.0/1540.0	1568.0
27	1542.5/1543.0	1563.0/1564.0	1530.0/1535.0	1515.0/1520.0	1543.0
28	1555.0/1555.5	1575.0/1576.0	1533.0/1538.0	1517.0/1522.0	1555.5
31	1542.5/1543.0	1564.0/1565.0	1523.0/1528.0	1508.0/1513.0	1543.0
2000/8/2	1552.0/1553.0	1575.0/1575.5	1533.0/1538.0	1515.0/1520.0	1553.0
3	1539.0/1540.0	1564.0/1565.0	1525.0/1530.0	1517.0/1522.0	1540.0
4	1523.0/1524.0	1549.5/1550.0	1518.0/1523.0	1508.0/1513.0	1524.0
7	1537.0/1538.0	1563.0/1563.5	1527.0/1532.0	1517.0/1522.0	1538.0
8	1534.0/1535.0	1558.0/1558.5	1528.0/1533.0	1518.0/1523.0	1535.0
9	1512.0/1513.0	1534.0/1535.0	1518.0/1523.0	1510.0/1515.0	1513.0
10	1505.5/1506.0	1530.0/1531.0	1515.0/1520.0	1512.0/1517.0	1506.0
11	1514.0/1515.0	1534.0/1535.0	1515.0/1520.0	1510.0/1515.0	1515.0
14	1515.0/1516.0	1537.5/1538.0	1515.0/1520.0	1510.0/1515.0	1516.0
15	1533.0/1533.5	1555.0/1555.5	1533.0/1538.0	1527.0/1532.0	1533.5
16	1521.5/1522.5	1545.5/1546.0	1525.0/1530.0	1520.0/1525.0	1522.5
17	1516.0/1517.0	1538.5/1539.0	1523.0/1528.0	1518.0/1523.0	1517.0
18	1521.0/1522.0	1547.0/1548.0	1532.0/1537.0	1525.0/1530.0	1522.0
21	1509.0/1509.5	1535.5/1536.0	1525.0/1530.0	1518.0/1523.0	1509.5
22	1509.0/1509.5	1538.0/1538.5	1530.0/1535.0	1523.0/1528.0	1509.5
23	1518.0/1518.5	1545.0/1545.5	1535.0/1540.0	1528.0/1533.0	1518.5

continued on next page

continued from previous page

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
24	1524.0/1525.0	1551.5/1552.0	1540.0/1545.0	1533.0/1538.0	1525.0
25	1521.0/1522.0	1550.0/1550.5	1533.0/1538.0	1525.0/1530.0	1522.0
28	***	***	***	***	***
29	1522.0/1522.5	1547.5/1548.0	1537.0/1542.0	1530.0/1535.0	1522.5
30	1555.0/1555.5	1583.0/1584.0	1560.0/1565.0	1547.0/1552.0	1555.5
31	1577.0/1578.0	1602.5/1603.5	1572.0/1577.0	1555.0/1560.0	1578.0
2000/9/1	1571.0/1572.0	1597.5/1598.0	1572.0/1577.0	1552.0/1557.0	1572.0
4	1562.5/1563.0	1588.0/1589.0	1565.0/1570.0	1545.0/1550.0	1563.0
5	1574.5/1575.0	1600.0/1600.5	1572.0/1577.0	1545.0/1550.0	1575.0
6	1609.0/1610.0	1631.0/1632.0	1583.0/1588.0	1545.0/1550.0	1610.0
7	1617.5/1618.0	1643.5/1644.0	1593.0/1598.0	1555.0/1560.0	1618.0
8	1635.0/1636.0	1654.0/1655.0	1593.0/1598.0	1553.0/1558.0	1636.0
11	1620.0/1620.5	1642.0/1642.5	1587.0/1592.0	1553.0/1558.0	1620.5
12	1616.0/1617.0	1639.5/1640.0	1580.0/1585.0	1555.0/1560.0	1617.0
13	1643.0/1644.0	1663.0/1663.5	1590.0/1595.0	1548.0/1553.0	1644.0
14	1638.0/1639.0	1657.0/1657.5	1588.0/1593.0	1545.0/1550.0	1639.0
15	1618.5/1619.0	1640.0/1640.5	1585.0/1590.0	1543.0/1548.0	1619.0
18	1602.0/1602.5	1621.5/1622.0	1568.0/1573.0	1533.0/1538.0	1602.5
19	1614.0/1615.0	1634.0/1634.5	1582.0/1587.0	1547.0/1552.0	1615.0
20	1610.5/1611.0	1629.0/1629.5	1572.0/1577.0	1537.0/1542.0	1611.0
21	1574.0/1574.5	1593.5/1594.0	1552.0/1557.0	1528.0/1533.0	1574.5

continued on next page

continued from previous page

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
22	1585.0/1586.0	1605.0/1606.0	1563.0/1568.0	1533.0/1538.0	1586.0
25	1587.0/1588.0	1603.0/1604.0	1560.0/1565.0	1530.0/1535.0	1588.0
26	1609.0/1609.5	1624.5/1625.0	1580.0/1585.0	1550.0/1555.0	1609.5
27	1570.0/1570.5	1591.5/1592.0	1553.0/1558.0	1535.0/1540.0	1570.5
28	1583.0/1584.0	1601.0/1601.5	1567.0/1572.0	1542.0/1547.0	1584.0
29	1578.0/1579.0	1595.0/1595.5	1565.0/1570.0	1540.0/1545.0	1579.0
2000/10/2	1573.0/1574.0	1588.5/1589.0	1558.0/1563.0	1532.0/1537.0	1574.0
3	1520.0/1520.5	1539.5/1540.0	1548.0/1553.0	1528.0/1533.0	1520.5
4	1515.5/1516.5	1534.5/1535.0	1538.0/1543.0	1520.0/1525.0	1516.5
5	1527.0/1528.0	1540.0/1541.0	1540.0/1545.0	1517.0/1522.0	1528.0
6	1527.0/1528.0	1536.0/1537.0	1533.0/1538.0	1512.0/1517.0	1528.0
9	1513.0/1513.5	1525.5/1526.0	1535.0/1540.0	1523.0/1528.0	1513.5
10	1517.0/1518.0	1529.0/1530.0	1540.0/1545.0	1530.0/1535.0	1518.0
11	1524.5/1525.5	1535.0/1536.0	1542.0/1547.0	1523.0/1528.0	1525.5
12	1531.0/1532.0	1538.0/1539.0	1532.0/1537.0	1513.0/1518.0	1532.0
13	1513.0/1513.5	1523.0/1523.5	1530.0/1535.0	1523.0/1528.0	1513.5
16	1508.0/1509.0	1525.0/1525.5	1532.0/1537.0	1520.0/1525.0	1509.0
17	1491.0/1492.0	1505.0/1505.5	1515.0/1520.0	1510.0/1515.0	1492.0
18	1481.0/1481.5	1500.0/1501.0	1512.0/1517.0	1513.0/1518.0	1481.5
19	1477.5/1478.0	1496.5/1497.0	1508.0/1513.0	1505.0/1510.0	1478.0
20	1489.0/1490.0	1503.5/1504.0	1512.0/1517.0	1505.0/1510.0	1490.0

continued on next page

continued from previous page

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
23	1485.0/1485.5	1500.5/15001.0	1513.0/1518.0	1505.0/1510.0	1485.5
24	1495.0/1496.0	1512.5/1513.0	1527.0/1532.0	1520.0/1525.0	1496.0
25	1480.0/1481.0	1497.5/1498.5	1520.0/1525.0	1515.0/1520.0	1481.0
26	1456.5/1457.0	1476.0/1477.0	1503.0/1508.0	1513.0/1518.0	1457.0
27	1445.0/1446.0	1466.0/1466.5	1498.0/1503.0	1510.0/1515.0	1446.0
30	1459.0/1460.0	1477.0/1478.0	1507.0/1512.0	1515.0/1520.0	1460.0
31	1468.0/1469.0	1485.0/1485.5	1512.0/1517.0	1520.0/1525.0	1469.0
2000/11/1	1477.5/1478.0	1494.5/1495.0	1518.0/1523.0	1523.0/1528.0	1478.0
2	1464.0/1465.0	1483.0/1483.5	1510.0/1515.0	1523.0/1528.0	1465.0
3	1465.0/1465.5	1482.0/1483.0	1508.0/1513.0	1515.0/1520.0	1465.5
6	1475.0/1476.0	1499.0/1500.0	1520.0/1525.0	1522.0/1527.0	1476.0
7	1472.5/1473.0	1496.5/1497.0	1520.0/1525.0	1522.0/1527.0	1473.0
8	1478.0/1478.5	1499.0/1500.0	1525.0/1530.0	1525.0/1530.0	1478.5
9	1474.0/1475.0	1498.0/1498.5	1513.0/1518.0	1510.0/1515.0	1475.0
10	1469.5/1470.0	1492.0/1493.0	1510.0/1515.0	1505.0/1510.0	1470.0
13	1451.0/1451.5	1473.0/1473.5	1503.0/1508.0	1508.0/1513.0	1451.5
14	1468.0/1469.0	1487.0/1487.5	1515.0/1520.0	1518.0/1523.0	1469.0
15	1472.0/1472.5	1489.0/1490.0	1515.0/1520.0	1517.0/1522.0	1472.5
16	1459.5/1460.5	1479.5/1480.5	1508.0/1513.0	1513.0/1518.0	1460.5
17	1445.0/1446.0	1469.0/1470.0	1503.0/1508.0	1515.0/1520.0	1446.0
20	1442.5/1443.0	1466.0/1467.0	1498.0/1503.0	1508.0/1513.0	1443.0

continued on next page

continued from previous page

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
21	1457.0/1457.5	1483.0/1483.5	1518.0/1523.0	1530.0/1535.0	1457.5
22	1470.0/1471.0	1495.0/1495.5	1525.0/1530.0	1528.0/1533.0	1471.0
23	1469.5/1470.0	1490.0/1491.0	1522.0/1527.0	1525.0/1530.0	1470.0
24	1483.0/1484.0	1502.5/1503.5	1530.0/1535.0	1533.0/1538.0	1484.0
27	1502.0/1502.5	1520.0/1521.0	1538.0/1543.0	1533.0/1538.0	1502.5
28	1512.0/1512.5	1529.0/1529.5	1540.0/1545.0	1532.0/1537.0	1512.5
29	1513.0/1514.0	1530.0/1531.0	1545.0/1550.0	1537.0/1542.0	1514.0
30	1497.0/1498.0	1513.5/1514.0	1533.0/1538.0	1528.0/1533.0	1498.0
2000/12/1	1493.0/1494.0	1508.5/1509.0	1530.0/1535.0	1525.0/1530.0	1494.0
4	1514.0/1515.0	1529.0/1529.5	1547.0/1552.0	1540.0/1545.0	1515.0
5	1505.0/1506.0	1520.0/1521.0	1537.0/1542.0	1530.0/1535.0	1506.0
6	1511.0/1512.0	1528.0/1529.0	1543.0/1548.0	1537.0/1542.0	1512.0
7	1548.0/1549.0	1557.0/1557.5	1560.0/1565.0	1565.0/1570.0	1549.0
8	1579.0/1580.0	1588.0/1589.0	1585.0/1590.0	1560.0/1565.0	1580.0
11	1631.5/1632.5	1631.0/1631.5	1593.0/1598.0	1553.0/1558.0	1632.5
12	1608.0/1609.0	1612.0/1613.0	1583.0/1588.0	1548.0/1553.0	1609.0
13	1587.0/1588.0	1597.0/1598.0	1567.0/1572.0	1540.0/1545.0	1588.0
14	1596.0/1597.0	1602.5/1603.0	1562.0/1567.0	1533.0/1538.0	1597.0
15	1620.0/1620.5	1622.0/1622.5	1580.0/1585.0	1552.0/1557.0	1620.5
18	1593.0/1594.5	1598.0/1599.0	1563.0/1568.0	1533.0/1538.0	1594.0
19	1601.5/1602.5	1603.0/1604.0	1570.0/1575.0	1540.0/1545.0	1602.5

continued on next page

continued from previous page

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
20	1574.0/1575.0	1572.5/1573.5	1547.0/1552.0	1520.0/1525.0	1575.0
21	1534.0/1534.5	1534.5/1535.5	1530.0/1535.0	1518.0/1523.0	1534.5
22	1558.0/1559.0	1556.0/1557.0	1548.0/1553.0	1533.0/1538.0	1559.0
25	***	***	***	***	***
26	***	***	***	***	***
27	1560.0/1560.5	1561.0/1561.5	1548.0/1553.0	1533.0/1538.0	1560.5
28	1562.0/1563.0	1558.0/1559.0	1545.0/1550.0	1532.0/1537.0	1563.0
29	1559.0/1560.0	1553.0/1554.0	1542.0/1547.0	1528.0/1533.0	1560.0
2001/1/2	1565.5/1566.5	1562.0/1562.5	1547.0/1552.0	1533.0/1538.0	1566.5
3	1520.0/1521.0	1524.0/1525.0	1518.0/1523.0	1508.0/1513.0	1521.0
4	1533.0/1533.5	1531.0/1532.0	1527.0/1532.0	1520.0/1525.0	1533.5
5	1538.0/1539.0	1529.0/1529.5	1523.0/1528.0	1520.0/1525.0	1539.0
8	1538.0/1539.0	1532.0/1533.0	1528.0/1533.0	1527.0/1532.0	1539.0
9	1580.0/1581.0	1566.0/1566.5	1550.0/1555.0	1532.0/1537.0	1581.0
10	1577.0/1578.0	1559.0/1560.0	1548.0/1553.0	1530.0/1535.0	1578.0
11	1575.5/1576.5	1556.5/1557.0	1538.0/1543.0	1520.0/1525.0	1576.5
12	1618.0/1620.0	1572.0/1573.0	1543.0/1548.0	1518.0/1523.0	1620.0
15	1629.0/1630.0	1567.0/1568.0	1535.0/1540.0	1512.0/1517.0	1630.0
16	1596.0/1597.0	1552.0/1553.0	1528.0/1533.0	1515.0/1520.0	1597.0
17	1575.5/1576.5	1547.0/1548.0	1533.0/1538.0	1520.0/1525.0	1576.5
18	1615.0/1616.0	1569.0/1570.0	1532.0/1537.0	1508.0/1513.0	1616.0

continued on next page

continued from previous page

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
19	1659.0/1660.0	1594.5/1595.5	1545.0/1550.0	1510.0/1515.0	1660.0
22	1650.0/1653.0	1595.5/1596.0	1545.0/1550.0	1513.0/1518.0	1653.0
23	1645.0/1646.0	1594.0/1594.5	1540.0/1545.0	1512.0/1517.0	1646.0
24	1652.0/1653.0	1605.0/1606.0	1542.0/1547.0	1517.0/1522.0	1653.0
25	1659.0/1660.0	1604.0/1604.5	1548.0/1553.0	1523.0/1528.0	1660.0
26	1683.0/1685.0	1618.0/1620.0	1563.0/1568.0	1537.0/1542.0	1685.0
29	1698.0/1700.0	1631.0/1632.0	1570.0/1575.0	1543.0/1548.0	1700.0
30	1688.0/1690.0	1621.0/1622.0	1562.0/1567.0	1535.0/1540.0	1690.0
31	1736.0/1737.0	1641.0/1642.0	1580.0/1585.0	1552.0/1557.0	1737.0
2001/2/1	1704.0/1705.0	1624.0/1624.5	1572.0/1577.0	1545.0/1550.0	1705.0
2	1690.0/1691.0	1608.0/1608.5	1563.0/1568.0	1535.0/1540.0	1691.0
5	1655.0/1655.5	1604.0/1605.0	1562.0/1567.0	1533.0/1538.0	1655.5
6	1625.0/1630.0	1603.0/1605.0	1565.0/1570.0	1540.0/1545.0	1630.0
7	1623.0/1625.0	1595.0/1596.0	1560.0/1565.0	1543.0/1548.0	1625.0
8	1624.0/1625.0	1594.0/1595.0	1555.0/1560.0	1540.0/1545.0	1625.0
9	1632.0/1633.0	1601.0/1602.0	1552.0/1557.0	1535.0/1540.0	1633.0
12	1624.0/1625.0	1589.0/1590.0	1552.0/1557.0	1540.0/1545.0	1607.0
13	1606.0/1607.0	1585.0/1585.5	1550.0/1555.0	1540.0/1545.0	1607.0
14	1592.0/1593.0	1577.5/1578.0	1547.0/1552.0	1540.0/1545.0	1593.0
15	1604.0/1604.5	1589.5/1590.0	1553.0/1558.0	1540.0/1545.0	1604.5
16	1616.0/1616.5	1598.0/1599.0	1555.0/1560.0	1542.0/1547.0	1616.5

continued on next page

continued from previous page

y/m/d	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
19	1592.0/1593.0	1591.0/1592.0	1548.0/1553.0	1530.0/1535.0	1593.0
20	1586.0/1587.0	1588.0/1589.0	1552.0/1557.0	1533.0/1538.0	1587.0
21	1572.0/1572.5	1571.5/1572.0	1537.0/1542.0	1520.0/1525.0	1572.5
22	1547.5/1548.0	1550.5/1551.0	1522.0/1527.0	1515.0/1520.0	1548.0

Table C.4: London metal exchange official closing prices of aluminium difference time series

Time	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
1	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.0
2	-13.5/-13.0	-9.0/-9.0	-5.0/-5.0	-1.0/-1.0	-13.0
3	0.5/0.5	2.5/3.0	0.0/0.0	-2.0/-2.0	0.5
4	8.5/8.0	11.0/10.5	8.0/8.0	5.0/5.0	8.0
5	-36.0/-35.5	-33.0/-33.0	-22.0/-22.0	-15.0/-15.0	-35.5
6	3.5/3.0	1.5/2.0	4.0/4.0	5.0/5.0	3.0
7	13.5/14.0	11.5/11.5	3.0/3.0	2.0/2.0	14.0
8	-6.0/-6.0	-1.0/-1.5	0.0/0.0	-4.0/-4.0	-6.0
9	15.0/15.0	11.5/11.5	8.0/8.0	7.0/7.0	15.0
10	10.0/9.5	8.0/8.0	2.0/2.0	-3.0/-3.0	9.5
11	-3.0/-2.5	-5.5/-5.5	-10.0/-10.0	-10.0/-10.0	-2.5
12	8.5/8.0	6.0/6.5	7.0/7.0	10.0/10.0	8.0
13	-2.0/-2.0	-1.0/-1.5	-5.0/-5.0	-7.0/-7.0	-2.0
14	-9.5/-9.5	-4.0/-4.0	0.0/0.0	5.0/5.0	-9.5
15	-12.5/-12.5	-7.5/-7.5	6.0/6.0	8.0/8.0	-12.5
16	-8.0/-7.5	-10.5/-10.0	-1.0/-1.0	5.0/5.0	-7.5
17	0.5/0.5	0.0/0.0	-2.0/-2.0	-3.0/-3.0	0.5
18	-11.0/-11.5	-8.0/-8.5	-3.0/-3.0	-2.0/-2.0	-11.5
19	8.0/8.0	5.0/5.0	3.0/3.0	5.0/5.0	8.0

continued on next page

continued from previous page

Time	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
20	-11.0/-11.0	-9.0/-9.0	-7.0/-7.0	-8.0/-8.0	-11.0
21	6.0/6.0	4.0/4.5	-1.0/-1.0	-3.0/-3.0	6.0
22	-38.5/-38.5	-39.0/-39.5	-27.0/-27.0	-19.0/-19.0	-38.5
23	-13.513.0	-13.0/-12.5	-5.0/-5.0	-13.0/-13.0	-13.0
24	-15.0/-15.0	-9.0/-9.0	0.0/0.0	7.0/7.0	-15.0
25	-14.5/-14.5	-20.0/-20.5	-15.0/-15.0	-2.0/-2.0	-14.5
26	0.5/0.0	1.0/1.5	3.0/3.0	0.0/0.0	0.0
27	-21.0/-20.5	-17.0/-17.0	-5.0/-5.0	0.0/0.0	-20.5
28	7.0/7.0	10.0/9.5	5.0/5.0	3.0/3.0	7.0
29	7.0/7.0	3.0/3.5	2.0/2.0	5.0/5.0	7.0
30	-9.0/-9.5	-9.0/-9.0	-13.0/-13.0	-15.0/-15.0	-9.5
31	0.0/0.5	1.0/1.0	3.0/3.0	0.0/0.0	0.5
32	-25.5/-25.5	-24.5/-24.5	-18.0/-18.0	-5.0/-5.0	-25.5
33	-42.5/-42.5	-44.5/-44.5	-22.0/-22.0	-15.0/-15.0	-42.5
34	30.0/30.0	31.0/31.0	25.0/25.0	24.0/24.0	30.0
35	7.0/7.0	8.0/8.0	10.0/10.0	11.0/11.0	7.0
36	-5.0/-5.0	-9.0/-9.0	-17.0/-17.0	-21.0/-21.0	-5.0
37	13.0/13.0	16.0/16.0	15.0/15.0	13.0/13.0	13.0
38	23.0/23.0	22.0/21.5	15.0/15.0	10.0/10.0	23.0
39	-13.0/-13.0	-15.5/-15.0	-20.0/-20.0	-22.0/-22.0	-13.0
40	3.0/3.0	4.5/4.5	4.0/4.0	4.0/4.0	3.0

continued on next page

continued from previous page

Time	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
41	-1.5/-1.5	-5.0/-5.5	-7.0/-7.0	-9.0/-9.0	-1.5
42	12.5/12.5	14.0/14.5	12.0/12.0	12.0/12.0	12.5
43	-24.0/-24.0	-25.0/-25.0	-15.0/-15.0	-10.0/-10.0	-24.0
44	19.0/18.5	20.0/20.0	16.0/16.0	18.0/18.0	18.5
45	-15.0/-15.0	-14.0/-14.5	-15.0/-15.0	-16.0/-16.0	-15.0
46	-10.0/-9.5	-9.5/-9.5	-1.0/-1.0	5.0/5.0	-9.5
47	23.0/22.5	21.5/22.0	16.0/16.0	13.0/13.0	22.5
48	-28.0/-28.0	-26.0/-26	-20.0/-20.0	-13.0/-13.0	-28.0
49	23.0/24.5	21.0/21.0	17.0/17.0	13.0/13.0	24.5
50	-5.0/-6.5	-8.0/-8.0	-12.0/-12.0	-15.0/-15.0	-6.5
51	42.0/42.5	38.5/38.0	30.0/30.0	28.0/28.0	42.5
52	24.0/23.5	22.5/22.5	10.0/10.0	5.0/5.0	23.5
53	-2.0/-1.5	-1.0/-0.5	0.0/0.0	2.0/2.0	-1.5
54	-14.5/-15.0	-14.5/-14.5	-11.0/-11.0	-10.0/-10.0	-15.0
55	-2.0/-2.0	0.0/0.0	-2.0/-2.0	32.0/32.0	-2.0
56	-3.0/-2.5	-2.5/-2.5	-3.0/-3.0	-39.0/-39.0	-2.5
57	-13.0/-13.5	-12.0/-12.0	-7.0/-7.0	-6.0/-6.0	-13.5
58	-11.5/-11.0	-7.0/-7.0	-7.0/-7.0	-7.0/-7.0	-11.0
59	-5.0/-5.0	-7.0/-7.0	0.0/0.0	5.0/5.0	-5.0
60	-7.5/-8.0	-4.5/-5.0	-5.0/-5.0	-5.0/-5.0	-8.0
61	-16.5/-16.0	-20.0/-20.0	-6.0/-6.0	2.0/2.0	-16.0

continued on next page

continued from previous page

Time	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
62	3.0/3.0	-0.5/0.0	-2.0/-2.0	-2.0/-2.0	3.0
63	21.0/21.0	11.5/11.0	3.0/3.0	-2.0/-2.0	21.0
64	-44.0/-44.5	-36.5/-36.0	-25.0/-25.0	-18.0/-18.0	-44.5
65	24.0/24.0	24.0/23.5	22.0/22.0	23.0/23.0	24.0
66	-6.0/-6.0	-6.0/-6.0	0.0/0.0	4.0/4.0	-6.0
67	6.0/6.5	5.5/5.5	8.0/8.0	8.0/8.0	6.5
68	13.0/12.5	11.5/12.0	2.0/2.0	-3.0/-3.0	12.5
69	1.0/1.5	2.0/1.5	-2.0/-2.0	-7.0/-7.0	1.5
70	-9.0/-9.0	-11.0/-11.0	-10.0/-10.0	-10.0/-10.0	-9.0
71	30.0/30.0	27.0/27.5	15.0/15.0	10.0/10.0	30.0
72	68.0/68.0	64.0/64.0	42.0/42.0	28.0/28.0	68.0
73	8.0/8.0	11.0/11.0	5.0/5.0	-1.0/-1.0	8.0
74	-4.0/-4.0	3.0/3.0	8.0/8.0	5.0/5.0	-4.0
75	-16.0/-16.0	-16.5/-17.0	-16.0/-16.0	-9.0/-9.0	-16.0
76	5.0/5.0	2.5/3.0	0.0/0.0	0.0/0.0	5.0
77	11.5/11.0	11.5/11.0	6.0/6.0	5.0/5.0	11.0
78	-6.5/-7.0	-7.0/-7.0	-10.0/-10.0	-13.0/-13.0	-6.5
79	10.0/11.0	16.0/16.0	17.0/17.0	17.0/17.0	10.5
80	25.0/24.5	21.5/22.0	18.0/18.0	16.0/16.0	24.5
81	-5.0/-4.5	-6.0/-6.0	-6.0/-6.0	-6.0/-6.0	-4.5
82	-18.0/-18.0	-17.0/-17.0	-15.0/-15.0	-14.0/-14.0	-18.0

continued on next page

continued from previous page

Time	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
83	7.0/7.0	5.0/5.0	6.0/6.0	5.0/5.0	7.0
84	-3.0/-3.5	-1.0/-1.5	-8.0/-8.0	-8.0/-8.0	-3.5
85	-8.07.5	-6.0/-5.5	-5.0/-5.0	-3.0/-3.0	-7.5
86	15.0/15.0	14.0/14.0	13.0/13.0	13.0/13.0	15.0
87	-7.07.5	-5.0/-5.0	-5.0/-5.0	-5.0/-5.0	-7.5
88	-18.0/-17.5	-16.0/-16.0	-10.0/-10.0	-10.0/-10.0	-17.5
89	-1.5/-2.0	-1.0/-1.0	0.0/0.0	0.0/0.0	-2.0
90	27.5/28.0	26.0/26.0	25.0/25.0	25.0/25.0	28.0
91	-5.5/-6.0	-5.0/-5.5	-8.0/-8.0	-7.0/-7.0	-6.0
92	10.0/10.5	4.5/5.0	-7.0/-7.0	-10.0/-10.0	-10.5
93	-11.5/-11.5	-9.5/-10.0	-10.0/-10.0	-10.0/-10.0	-11.5
94	10.5/10.5	9.5/9.5	5.0/5.0	7.0/7.0	10.5
95	26.5/26.5	26.0/26.0	15.0/15.0	13.0/13.0	26.5
96	-12.0/-12.0	-10.5/-10.0	-10.0/-10.0	-10.0/-10.0	-12.0
97	-24.0/-24.5	-22.0/-22.5	-13.0/-13.0	-13.0/-13.0	-24.5
98	18.5/18.5	16.0/16.0	7.0/7.0	5.0/5.0	18.5
99	-12.5/-12.0	-12.0/-12.0	-5.0/-5.0	-3.0/-3.0	-12.0
100	-1.0/-1.5	-1.0/-0.5	-7.0/-7.0	-5.0/-5.0	-1.5
101	0.0/0.5	0.5/0.0	5.0/5.0	3.0/3.0	0.5
102	-24.5/-25.0	-25.5/-25.0	-20.0/-20.0	-20.0/-20.0	-25.0
103	12.5/12.5	12.0/12.0	3.0/3.0	2.0/2.0	12.5

continued on next page

continued from previous page

Time	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
104	-12.5/-12.5	-11.0/-11.0	-10.0/-10.0	-9.0/-9.0	-12.5
105	9.5/10.0	11.0/10.5	10.0/10.0	7.0/7.0	10.0
106	-13.0/-13.0	-11.0/-10.5	-8.0/-8.0	2.0/2.0	-13.0
107	-16.0/-16.0	-14.5/-15.0	-7.0/-7.0	-9.0/-9.0	-16.0
108	14.0/14.0	13.5/13.5	9.0/9.0	9.0/9.0	14.0
109	-3.0/-3.0	-5.0/-5.0	1.0/1.0	1.0/1.0	-3.0
110	-22.0/-22.0	-24.0/-23.5	-10.0/-10.0	-8.0/-8.0	-22.0
111	-6.5/-7.0	-4.0/-4.0	-3.0/-3.0	2.0/2.0	-7.0
112	8.5/9.0	4.0/4.0	0.0/0.0	-2.0/-2.0	9.0
113	1.5/1.0	3.5/3.0	0.0/0.0	0.0/0.0	1.0
114	17.5/17.5	17.5/17.5	18.0/18.0	17.0/17.0	17.5
115	-11.5/-11.0	-9.5/-9.5	-8.0/-8.0	-7.0/-7.0	-11.0
116	-5.5/-5.5	-7.0/-7.0	-2.0/-2.0	-2.0/-2.0	-5.5
117	5.0/5.0	8.5/9.0	9.0/9.0	7.0/7.0	5.0
118	-12.0/-12.5	-11.5/-12.0	-7.0/-7.0	-7.0/-7.0	-12.5
119	0.0/0.0	2.5/2.5	5.0/5.0	5.0/5.0	0.0
120	9.0/9.0	7.0/7.0	5.0/5.0	5.0/5.0	9.0
121	6.0/6.5	6.5/6.5	5.0/5.0	5.0/5.0	6.5
122	-3.0/-3.0	-1.5/-1.5	-7.0/-7.0	-8.0/-8.0	-3.0
123	1.0/0.5	-2.5/-2.5	4.0/4.0	5.0/5.0	0.5
124	33.0/33.0	35.5/36.0	23.0/23.0	17.0/17.0	33.0

continued on next page

continued from previous page

Time	Cash	3 months	15 months	27 months	Sett
	Bid/Ask	Bid/Ask	Bid/Ask	Bid/Ask	
125	22.0/22.5	19.5/19.5	12.0/12.0	8.0/8.0	22.5
126	-6.0/-6.0	-5.0/-5.5	0.0/0.0	-3.0/-3.0	-6.0
127	-8.5/-9.0	-9.5/-9.0	-7.0/-7.0	-7.0/-7.0	-9.0
128	12.0/12.0	12.0/11.5	7.0/7.0	0.0/0.0	12.0
129	34.5/35.0	31.0/31.5	11.0/11.0	0.0/0.0	35.0
130	8.5/8.0	12.5/12.0	10.0/10.0	10.0/10.0	8.0
131	17.5/18.0	10.5/11.0	0.0/0.0	-2.0/-2.0	18.0
132	-15.0/-15.5	-12.0/-12.5	-6.0/-6.0	0.0/0.0	-15.5
133	-4.0/-3.5	-2.5/-2.5	-7.0/-7.0	2.0/2.0	-3.5
134	27.0/27.0	23.5/23.5	10.0/10.0	-7.0/-7.0	27.0
135	-5.0/-5.0	-6.0/-6.0	-2.0/-2.0	-3.0/-3.0	-5.0

C.3 Table of Chapter 7, Section 7.5

Table C.5: Bivariate simulated series

time	Series 1	Series 2
1	2.032	2.139
2	3.816	-1.053
3	4.068	-1.513
4	7.799	4.969
5	11.143	4.334
6	10.429	3.523
7	12.443	3.637
8	13.437	5.193
9	13.273	2.321
10	14.958	4.229
11	19.056	5.828
12	21.452	4.540
13	22.318	7.646
14	24.673	7.155
15	28.066	6.481
16	30.246	7.492
17	30.811	9.645
18	30.763	8.512
19	30.748	8.742
20	25.594	9.929

continued on next page

continued from previous page

time	Series 1	Series 2
21	27.199	8.078
22	27.969	5.586
23	23.158	6.793
24	24.847	4.916
25	28.130	6.844
26	30.454	6.212
27	30.873	8.639
28	32.263	10.974
29	35.084	10.289
30	39.607	13.623
31	43.796	14.020
32	48.920	15.763
33	49.445	14.027
34	57.814	14.799
35	57.726	16.265
36	64.836	20.110
37	68.117	18.690
38	74.463	21.967
39	77.167	24.339
40	83.118	24.372
41	91.167	25.606
42	96.721	27.561

continued on next page

continued from previous page

time	Series 1	Series 2
43	108.881	30.045
44	109.296	31.272
45	114.313	31.552
46	119.979	34.740
47	125.932	36.445
48	132.849	39.508
49	143.335	45.071
50	150.244	46.284
51	156.628	45.250
52	156.919	46.812
53	174.764	52.067
54	173.397	51.018
55	181.088	51.075
56	190.850	56.286
57	192.046	57.955
58	204.229	63.790
59	205.190	60.668
60	206.672	62.032
61	212.163	63.050
62	220.828	65.489
63	226.267	69.658
64	231.180	66.638

continued on next page

continued from previous page

time	Series 1	Series 2
65	233.643	69.435
66	237.435	71.024
67	244.024	78.131
68	247.346	76.193
69	244.893	75.056
70	257.685	74.807
71	265.736	83.327
72	265.897	77.735
73	271.066	81.055
74	271.996	84.209
75	281.346	84.950
76	291.439	87.386
77	297.085	87.001
78	308.921	90.093
79	306.365	95.398
80	315.729	96.805
81	322.246	94.436
82	333.993	98.402
83	336.767	102.933
84	342.449	105.093
85	351.786	101.286
86	361.939	114.032

continued on next page

continued from previous page

time	Series 1	Series 2
87	375.531	106.811
88	381.458	114.362
89	390.112	114.547
90	398.871	117.000
91	406.865	124.967
92	412.168	126.189
93	426.345	128.359
94	433.726	128.834
95	442.306	132.683
96	452.102	135.543
97	462.907	137.692
98	469.279	137.761
99	481.880	144.591
100	487.071	143.287
101	496.037	146.109
102	504.166	149.255
103	511.015	153.888
104	522.588	156.207
105	531.781	161.039
106	541.111	161.161
107	547.769	160.619
108	555.831	163.937

continued on next page

continued from previous page

time	Series 1	Series 2
109	564.021	167.085
110	570.758	168.853
111	587.673	170.655
112	593.369	176.003
113	591.953	176.258
114	607.279	181.743
115	611.448	183.276
116	616.653	186.233
117	634.774	189.469
118	641.892	190.671
119	647.075	194.133
120	663.107	195.825
121	677.814	202.293
122	684.712	206.883
123	697.299	206.000
124	708.245	212.654
125	720.715	213.045
126	725.218	219.879
127	737.188	219.243
128	747.358	223.435
129	758.266	225.027
130	761.789	227.481

continued on next page

continued from previous page		
time	Series 1	Series 2
131	777.798	232.504
132	782.626	234.523
133	790.877	238.345
134	798.031	236.610
135	803.873	239.203
136	812.336	241.095
137	824.572	246.298
138	828.278	248.454
139	837.210	250.601
140	851.259	252.219
141	856.838	255.971
142	866.560	255.730
143	868.641	258.445
144	879.938	261.635
145	889.710	265.970
146	893.876	267.270
147	900.348	267.415
148	907.970	271.681
149	913.355	272.438
150	921.531	275.049

Bibliography

[1] J.R.M. Ameen. *Discount Bayesian models and forecasting.* PhD thesis, University of Warwick, 1984.

[2] J.R.M. Ameen and P.J. Harrison. Normal discount Bayesian models. In *Bayesian Statistics 2*, J.M. Bernanrdo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith (Eds.). North-Holland, Amsderdam, and Valencia University Press, 1985.

[3] E. Barbosa. *Dynamic Bayesian models for vector time series analysis.* PhD thesis, University of Warwick, 1989.

[4] E. Barbosa and P.J. Harrison. Variance estimation for multivariate DLMs. *J. Forecasting*, 11:621–628, 1992.

[5] G.E.P. Box and G.C. Tiao. *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Massachusetts, 1970.

[6] P.J. Brown, N.D. Le, and J.V. Zidek. Inference for a covariance matrix. In *Aspects of Uncertainty: a tribute to D.V. Lindley*, P.R. Freeman and A.F.M. Smith (Eds.). Wiley, New York, 1994.

[7] C.K. Carter and R. Kohn. On Gibbs sampling for state space models. *Biometrika*, 81:541–543, 1994.

[8] A.P. Dawid. Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika*, 68:265–274, 1981.

[9] A.P. Dawid and S.L. Lauritzen. Hyper Markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics*, 21:1272–1317, 1993.

[10] J. Durbin and S.J. Koopman. *Time Series Analysis by State Space Methods.* Oxford Statistical Science Series, Oxford, 2001.

[11] E.F. Fama. Efficient capital markets: a review of theory and empirical work (with discussion). *J. Finance*, 25:383–423, 1970.

[12] S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *J. Time Series Analysis*, 14:183–202, 1994.

[13] P.H. Garthwaite and S.A. Al-Awadhi. Non-conjugate prior distribution assessment for multivariate normal sampling. *J. Roy. Statist. Soc. (Ser B)*, 63:95–110, 2001.

[14] A.K. Gupta and D.K. Nagar. *Matrix Variate Distributions.* Chapman and Hall, New York, 1999.

[15] P.J. Harrison. Short-term sales forecasting. *Applied Statistics*, 15:102–139, 1965.

[16] P.J. Harrison. Exponential smoothing and short-term forecasting. *Management Science*, 13:821–842, 1967.

[17] P.J. Harrison. Convergence and the constant dynamic linear model. *J. Forecasting*, 16:287–292, 1997.

[18] P.J. Harrison and C. Stevens. A Bayesian approach to short-term forecasting. *Operations Research Quarterly*, 22:341–362, 1971.

[19] P.J. Harrison and C. Stevens. Bayesian forecasting (with dicussion). *J. Roy. Statist. Soc. (Ser B)*, 38:205–247, 1976.

[20] P.J. Harrison and P.P. Veerapen. Incorporating and deleting information in dynamic models. In *Developments in Time Series Analysis: in Honour of Maurice B. Priestley*, T. Subba Rao (Ed.). Chapman and Hall, London, 1993.

[21] P.J. Harrison and M. West. Bayesian forecasting in practice. *Bayesian Statistics Study Year Report* 13, University of Warwick, 1986.

[22] P.J. Harrison and M. West. Practical Bayesian forecasting. *The Statistician*, 36:115–125, 1987.

[23] P.J. Harrison and M. West. Dynamic linear model diagnostics. *Biometrika*, 78:797–808, 1991.

[24] D.A. Harville. *Matrix Algebra from a Statistician's Perspective.* Springer-Verlag, 1997.

[25] R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, 1999.

[26] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:05–45, 1960.

[27] R.E. Kalman. New methods in Wiener filtering theory. In *Proceedings of the First Symposium of Engineering Applications of Random Function Theory and Probability*, J.L. Bogdanoff and F. Kozin (Eds.). Wiley, New York. 1963.

[28] K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis.* Academic Press, London, 1979.

[29] R.J. Muirhead. *Aspects of Multivariate Statistical Theory.* Wiley, New York, 1982.

[30] A. Pole and M. West. Reference analysis of the DLM. *Time Series Analysis*, 10:131–147, 1989.

[31] A. Pole, M. West, and P.J. Harrison. *Applied Bayesian Forecasting and Time Series Analysis.* Chapman and Hall, New York, 1994.

[32] S.J. Press. *Applied Multivariate Analysis.* Holt, Rinehart and Winston, 1972.

[33] J.M. Quintana. A dynamic linear matrix-variate regression model. *Research Report* 83, Department of Statistics, University of Warwick, 1985.

[34] J.M. Quintana. *Multivariate Bayesian forecasting models.* PhD thesis, University of Warwick, 1987.

[35] J.M. Quintana, V.K. Chopra, and B.H. Putnam. Global asset allocation: Stretcling returns by shrinking forecasts. In *Proceedings of the ASA Section on Bayesian Statistical Science, 1995 Joint Statistical Meetings*, American Statkstical Association, 1995.

[36] J.M. Quintana and M. West. Multivariate time series analysis: new techniques applied to international exchange rate data. *The Statistician*, 88:275–281, 1987.

[37] J.M. Quintana and M. West. Time series analysis of compositional data. In *Bayesian Statistics 3*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith (Eds.). Oxford University Press. 1988.

[38] A. Roverato. Cholesky decomposition of a hyper inverse Wishart matrix. *Biometrika*, 87.

[39] R.H. Shumway and D.S. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3:253–264, 1982.

[40] K. Triantafyllopoulos. Missing observations in multivariate dynamic linear models. *Research Report* 376, Department of Statistics, University of Warwick, 2000.

[41] K. Triantafyllopoulos. Stochastic variance laws in multivariate DLMs. *Research Report* 383, Department of Statistics, University of Warwick, 2001.

[42] K. Triantafyllopoulos and P.J. Harrison. Weak probability and dynamic linear models. *Research Report* 379, Department of Statistics, University of Warwick, 2001.

[43] K. Triantafyllopoulos and J. Pikoulas. Bayesian multivariate regression for predicting user behaviour in a software agent computer security environment. *Research Report* 375, Department of Statistics, University of Warwick, 2000.

[44] K. Triantafyllopoulos, J. Pikoulas, and W.J. Buchanan. An intelligent intrusion detection environment using software agents. In *Proceedings of the 13th International Conference of Software and Systems Engineering and Applications*, Vol. 4, Session 15.2, CNAM Paris 6–8 December 2000.

[45] K. Triantafyllopoulos, J. Pikoulas, W.J. Buchanan, and M. Manion. An agent based Bayesian forecasting model for enhanced network security. In *Proceedings of the 8th IEEE International Conference and Workshop on the Engineering of Computer Based Systems–ECBS*, IEEE Comput. Soc., 247–254, Los Alamitos, CA, 17–20 April 2001.

[46] H. Uligh. On singular Wishart and singular multivariate beta distributions. *Ann. Statist*, 23:395–405, 1994.

[47] H. Uligh. Bayesian vector-autoregressions with stochastic volatility. *Econometrica*, 65:59–63, 1997.

[48] P.P. Veerapen. *Recurrence relationships and model monitoring for dynamic linear models.* PhD thesis, University of Warwick, 1991.

[49] M. West. *Aspects of recursive Bayesian estimation.* PhD thesis, University of Nottingham, 1982.

[50] M. West. Bayesian forecasting. In *Encyclopedia of Statistical Sciences*, S. Kotz, C.B. Read, and D.L. Banks (Eds.). Wiley, New York. 1989.

[51] M. West and P.J. Harrison. *Bayesian Forecasting and Dynamic Models.* Springer-Verlag, New York, second edition, 1997.

Άνθη της πέτρας μπροστά στην πράσινη θάλασσα
με φλέβες που μου θύμιζαν άλλες αγάπες
γυαλίζοντας στ' αργό ψιχάλισμα,
άνθη της πέτρας φυσιογνωμίες
που ήρθαν όταν κανένας δε μιλούσε και μου μίλησαν
που μ' άφησαν να τις αγγίξω ύστερ' απ' τη σιωπή
μέσα σε πεύκα σε πικροδάφνες και σε πλατάνια.

Σχέδια Για Ένα Καλοκαίρι
Γιώργος Σεφέρης